

UNIVERSIDADE DE LISBOA

FACULDADE DE LETRAS



**ANÁLISE DA TRADUÇÃO DE  
COMBINATÓRIAS LEXICAIS EM TAREFAS DE  
PÓS-EDIÇÃO DE TRADUÇÃO AUTOMÁTICA**

TATIANE ROCHA OLIVEIRA

Dissertação orientada pela Professora Doutora Madalena Colaço e  
coorientada pela Professora Doutora Helena Moniz, especialmente  
elaborada para a obtenção do grau de Mestre em TRADUÇÃO

2019

## Resumo

A presente dissertação teve como base a análise de dados obtidos a partir de um estágio realizado na *start-up* Unbabel e trata da tradução de combinatórias lexicais de inglês para português do Brasil na área de tradução automática e subsequente pós-edição. As combinatórias lexicais são expressões complexas que, por suas propriedades, causam problemas de tradução especiais, sobretudo para a tradução automática que, como toda área tecnológica, está em constante desenvolvimento. Presentes em todos tipos de comunicação e em todas línguas, tanto na língua comum quanto na linguagem especializada, estas expressões são realmente relevantes e o seu estudo indispensável. A pesquisa analisou um *corpus* proveniente do processo de anotação da empresa. Observamos as combinatórias lexicais tanto em relação à anotação dos erros, quanto à identificação das origens dos mesmos. Os resultados mais significantes demonstraram a necessidade de incluir um tipo de erros para as combinatórias lexicais na taxonomia da empresa, bem como critérios de anotação para tais. A análise também apontou para a falha na detecção de variedades linguísticas em relação às combinatórias lexicais. Desta forma, também criamos um glossário de combinatórias lexicais que, além de conter as soluções encontradas para o português do Brasil, inclui os diferentes equivalentes para português europeu, contribuindo também para a distinção entre as duas variedades linguísticas em relação a estas expressões, uma vez que os estudos comparativos entre ambas tendem a tratar fundamentalmente das diferenças gramaticais.

**PALAVRAS-CHAVE:** combinatórias lexicais, tradução, pós-edição, anotação, português

## Abstract

This thesis was based on the analysis of data obtained from an internship in the AI Customer Service start-up Unbabel. The goal is to analyse and discuss the translation of lexical combinations from English into Brazilian Portuguese in the area of machine translation and subsequent post-editing. Lexical combinations are complex expressions that, due to their properties, cause many problems of translation, especially in machine translation. Present in all types of communication and in all languages, both in the general and in the specialized language, these expressions are very relevant, and their study is indispensable. Through analysis of a *corpus* from the company's annotation process, I observed lexical combinations related both to the annotation of errors and to the identification of their source. The most significant results demonstrated the need to include a type of error for lexical combinations in the company's taxonomy, as well as the annotation criteria for such. The research also highlighted the lack of language variety detection regarding lexical combinations. Therefore, I created a glossary of lexical combinations that, in addition to containing the equivalent solutions for Brazilian Portuguese, includes the different equivalents in European Portuguese, furthermore contributing to the distinction between the two language varieties regarding these expressions, as the comparative studies between them tend to deal essentially with the grammatical differences.

*Keywords:* lexical combinations, translation, post-editing, annotation, Portuguese

## **Agradecimentos**

À Professora Doutora Madalena Colaço por toda atenção, paciência e profissionalismo na orientação deste trabalho.

À Professora Doutora Helena Moniz pela supervisão na empresa, bem como por toda motivação e conselhos dados na coorientação da dissertação.

À Unbabel, pela oportunidade de fazer este estágio.

À Professora Doutora Guilhermina Jorge, pela inspiração no campo das Fraseologias.

À minha família, por me incentivar a realizar os meus sonhos.

Ao meu amor, Richard, por ter me acompanhado e me apoiado em todos os momentos desta jornada e à sua querida família, que se tornou minha também.

A todos os meus amigos do Brasil, da Irlanda, de Portugal e do mundo todo, que tornaram esta experiência possível e mais divertida, especialmente, as amigas de mais de uma década, a Vanessa e a Erika.

Às amigas parceiras do estágio e da FLUL, Nana, Rhandra e Catarina, e ao grupo de estudos formado também pela Nana, juntamente com a Joana, o Nuno e, posteriormente, a Priscila.

Gratidão!

## Sumário

1	Introdução .....	1
1.1	Objetivos .....	3
2	A entidade de acolhimento do estágio .....	5
3	Estado da arte.....	9
3.1	Características das combinatórias lexicais .....	9
3.1.1	Fixidez.....	13
3.1.2	Recorrência .....	14
3.1.3	Sentido figurado.....	15
3.2	Tipologia das combinatórias lexicais .....	15
3.2.1	Colocações .....	24
3.2.1.1	Transparência semântica.....	25
3.2.1.2	O papel da restrição combinatória .....	26
3.2.1.3	Tipos de colocações.....	29
3.2.2	Fraseologias .....	32
3.2.3	Expressões Idiomáticas .....	33
3.3	Problemas de tradução das combinatórias lexicais .....	37
3.3.1	Combinatórias lexicais e Tradução Automática .....	43
3.3.1.1	Sistemas de Tradução Automática .....	43
3.3.1.2	Desafios das combinatórias lexicais para a tradução automática .....	46
3.3.2	Combinatórias lexicais especializadas .....	50
3.3.3	Combinatórias lexicais e variedades linguísticas.....	53
3.3.3.1	Português do Brasil e português europeu .....	55
3.3.3.2	Diferenças entre português do Brasil e português europeu aplicadas às combinatórias lexicais.....	57
3.3.3.3	Diferenças entre português do Brasil e português europeu nas combinatórias lexicais da informática .....	68
3.4	Estratégias de tradução das combinatórias lexicais.....	69
4	Metodologia.....	74
4.1	Processo de anotação.....	74
4.1.1	Tipologia .....	75
4.1.2	Métrica .....	77
4.1.3	Critérios de anotação.....	79
4.2	Descrição do <i>corpus</i> .....	81
4.3	Criação de glossário .....	84
5	Análise dos dados .....	85

5.1	Identificação dos domínios.....	86
5.2	A anotação dos erros .....	91
5.2.1	Frequência dos erros de combinatórias lexicais.....	91
5.2.2	Anotação dos erros de combinatórias lexicais .....	92
5.3	Erros de tradução de combinatórias lexicais .....	104
5.3.1	Descrição dos erros .....	105
5.3.2	Explicação dos erros .....	111
5.4	Sugestões para a empresa.....	116
6	Conclusões.....	118
7	Referências Bibliográficas.....	120
	Anexo - Glossário .....	1

## Lista de tabelas

TABELA 1 DIFERENÇAS ENTRE UNIDADES FRASEOLÓGICAS DA LÍNGUA GERAL E UNIDADES FRASEOLÓGICAS ESPECIALIZADAS (EXTRAÍDO DE LORENTE, 2001, p. 19).....	12
TABELA 2 COLLOCATIONS VS. IDIOMS AND FREE WORD COMBINATIONS. (EXTRAÍDO DE MCKEOWN & RADEV, 2000, p. 3) .....	23
TABELA 3 SYNTACTIC PATTERNS OF COLLOCATION (EXTRAÍDO DE POULSEN, 2005, p. 14) .....	30
TABELA 4 MAGN (EXTRAÍDO DE VALENTE, 2000).....	60
TABELA 5 BON (EXTRAÍDO DE VALENTE, 2000) .....	60
TABELA 6 ANTIVER (EXTRAÍDO DE VALENTE, 2000).....	61
TABELA 7 OPER (EXTRAÍDO DE VALENTE, 2000) .....	62
TABELA 8 REAL (EXTRAÍDO DE VALENTE, 2000).....	62
TABELA 9 QUANTIDADE DE SEGMENTOS RECEBIDOS .....	82
TABELA 10 EXEMPLO DE AMBIGUIDADE .....	86
TABELA 11 FINANÇAS.....	89
TABELA 12 DOCUMENTAÇÃO.....	89
TABELA 13 COMUNICAÇÕES E COMÉRCIO .....	90
TABELA 14 TRANSPORTE E ENTRETENIMENTO .....	90
TABELA 15 QUANTIDADE DE ERROS DE TRADUÇÃO DE COMBINATÓRIAS LEXICAIS POR TIPO DE ERRO.....	91
TABELA 16 COMPARAÇÃO ENTRE AS ANOTAÇÕES DO CORPUS E AS ANOTAÇÕES DE ERROS DE COMBINATÓRIAS LEXICAIS.....	93
TABELA 17 INCONSISTÊNCIA NA ANOTAÇÃO DAS UNIDADES LEXICAIS.....	94
TABELA 18 INCONSISTÊNCIA NA ANOTAÇÃO DE EXPRESSÕES.....	97
TABELA 19 INCONSISTÊNCIA NA ANOTAÇÃO DOS TIPOS DE ERRO .....	98
TABELA 20 SEGMENTOS COM MAIS DE UMA ANOTAÇÃO .....	99
TABELA 21 COMO OS ERROS DE COMBINATÓRIAS LEXICAIS FORAM ANOTADOS .....	100
TABELA 22 DISTÂNCIA COLOCACIONAL .....	101
TABELA 23 EXEMPLO DE MESMO SEGMENTO ANOTADO TRÊS VEZES .....	102
TABELA 24 EXEMPLO DE ANOTAÇÃO PARCIAL E TIPO DE ERRO INCORRETO.....	103
TABELA 25 ERROS DE SUBSTITUIÇÃO .....	106
TABELA 26 OPÇÃO MELHOR .....	107
TABELA 27 SENTIDO DIFERENTE.....	108
TABELA 28 CRIAÇÃO .....	109
TABELA 29 ANÁLISE .....	109
TABELA 30 PREPOSIÇÃO INCORRETA.....	110
TABELA 31 SÍNTESE .....	110
TABELA 32 IMPORTAÇÃO .....	113
TABELA 33 VARIEDADE LINGUÍSTICA.....	113
TABELA 34 ESCOLHA ERRADA.....	114
TABELA 35 EXTENSÃO .....	114
TABELA 36 SOBREGENERALIZAÇÃO.....	115
TABELA 37 DERIVAÇÃO ERRADA .....	115

## Lista de figuras

FIGURA 1: UNBABEL'S TRANSLATION PIPELINE. (EXTRAÍDO DE UNBABEL, 2018B, P.3) .....	7
FIGURA 2: CLASSIFICAÇÃO DE COMBINAÇÕES DE PALAVRAS DE HAUSMANN (1984 APUD COSTA 2017, P. 19). .....	18
FIGURA 3: CLASSIFICAÇÃO DE COMBINAÇÕES DE PALAVRAS DE MEL'ČUK (1998) .....	19
FIGURA 4: CLASSIFICAÇÃO DE COMBINAÇÕES DE PALAVRAS DE COWIE (2001[1998] APUD COSTA, 2017, P. 22) .....	20
FIGURA 5: CLASSIFICAÇÃO DE COMBINAÇÕES DE PALAVRAS DE BURGER (2003 APUD COSTA, 2017, P. 23).....	21
FIGURA 6. TIPOLOGIA DE ERROS DO SISTEMA DE ANOTAÇÃO. (EXTRAÍDO DE UNBABEL, 2018B, P.15) .....	75
FIGURA 7: THREE SEVERITY LEVELS AND INTERSECTIONS FOR TYPE OF ERRORS. (EXTRAÍDO DE UNBABEL, 2018B, P. 24) .....	78
FIGURA 8: DOMÍNIOS.....	88
FIGURA 9: DESCRIÇÃO DOS ERROS .....	106
FIGURA 10: EXPLICAÇÃO DOS ERROS .....	112



## Lista de siglas

CL	combinatória lexical
CLE	combinatória lexical especializada
EI	expressão idiomática
FL	função lexical
HQ	<i>human-quality</i>
LC	língua de chegada
LP	língua de partida
MQM	<i>Multidimensional Quality Metrics</i>
MT	<i>Machine Translation</i>
NMT	<i>Neural Machine Translation</i>
PB	português do Brasil
PE	português europeu
QE	<i>Quality Estimation</i>
RBMT	<i>Ruled based Machine Translation</i>
RGPD	Regulamento Geral da Proteção de Dados
SMT	<i>Statistical Machine Translation</i>
TA	tradução automática
TC	texto de chegada
TI	Tecnologia da Informação
TP	texto de partida
TST	Teoria Sentido-texto
UFE	unidade fraseológica especializada

# 1 Introdução

O presente trabalho teve como base a análise de um conjunto de dados obtidos num estágio realizado na empresa “Unbabel” no ano letivo de 2017/2018. Foi redigido em português do Brasil (PB) e de acordo com o novo acordo ortográfico da língua portuguesa.

O estágio consistia na análise de resultados de tradução automática (TA) e de TA revista por revisores humanos da empresa, com vista à identificação de regularidades e à definição de especificações linguísticas com potencial para melhorar o desempenho da tecnologia de tradução desenvolvida pela empresa.

Assim, propomos uma reflexão sobre a tradução das combinatórias lexicais (CLs). De forma geral, os estudos sobre o léxico da língua portuguesa (independentemente da variedade) são ainda muito limitados no que diz respeito às CLs no âmbito da TA. Estas expressões existem em todas as línguas e em todos os tipos de comunicação, inclusive, alguns tipos de CLs ocorrem tanto na língua geral quanto na linguagem especializada.

Vilela (2002) defende que as CLs são ainda mais frequentes em textos predominantemente comunicativos e onde a oralidade predomina. Por esta razão, para a Unbabel, essas expressões são altamente relevantes, visto que a maior parte do material que a empresa traduz é de conteúdo dinâmico, como e-mails, tíquetes de suporte, legendas, etc., ou seja, repleta de marcas da oralidade.

Estas expressões são complexas e difíceis de definir ou delimitar. Alguns autores optam por estabelecer as fronteiras e fazer a distinção entre os diferentes tipos, enquanto outros autores optam por tratá-las conjuntamente, como um grupo de CLs. No presente trabalho, optamos por não fazer a distinção, partindo do princípio de que, mais fixas ou menos fixas, os problemas que as CLs causam para a tradução são os mesmos.

O estudo das CLs tem distintos propósitos, tais como: i) criação de dicionários, ii) ensino/aprendizagem das línguas estrangeiras e iii) tradução. Costa (2017), por exemplo, foca

na análise de colocações para o ensino de português como segunda língua e faz uma comparação de erros de alunos aprendentes com erros de TA. O presente trabalho foca nos problemas que as CLs causam para a tradução, especificamente para a TA e sua pós-edição humana. Para isso, optamos por analisar o processo de anotação da empresa, pois este processo possibilita a identificação e categorização dos erros tradução, além de ser crucial para avaliar a qualidade. Assim, observamos o processo especificamente em relação à anotação dos erros de tradução de CLs, de inglês para PB, identificamos as origens destes erros e criamos um glossário com as sugestões de tradução das CLs. A finalidade é prover uma visão abrangente sobre o fenômeno, explorar os vários problemas que estas expressões envolvem (e não um único problema, como focam vários autores), e apresentar as soluções encontradas. Afinal, uma visão geral dos potenciais problemas possibilita mais opções de resolução dos mesmos para a empresa, considerando os vários fatores envolvidos. De um modo geral, esta dissertação apresenta o estado da arte das CLs e verifica nos resultados da empresa os problemas e questões levantados, indicando as possíveis soluções.

Sinclair (1991) e McKeown & Radev (2000) defendem que a propriedade “recorrência” das CLs é definitiva para a TA. Os estudos de Grossmann e Tutin (2002), Mel’čuk (1998) e Hausmann (1984), contudo, comprovaram que apenas os dados estatísticos não são suficientes para a identificação das CLs, mas apenas complementares a outras análises.

Jorge (1997a) mostra alguns problemas de tradução das CLs. Por serem idiossincráticas, verificamos que um dos impactos são as diferenças entre variedades linguísticas. Neste sentido, comparamos as variedades “português europeu (PE) e PB” pois, na literatura, muito já se falou das diferenças gramaticais destas variedades, mas no que diz respeito às CLs, os trabalhos contrastivos são ainda escassos. Deste modo, no presente trabalho, contribuímos também neste aspecto.

Fundamental para esta comparação, Ortiz Alvarez (2012) contribui com o estado da arte da fraseologia do Brasil. Camacho (2008) e Valente (2000) apresentam algumas análises contrastivas das CLs nas duas variedades. No entanto, diferente do presente trabalho, as autoras fazem o estudo comparativo em tipos específicos de CLs (Camacho trata das expressões idiomáticas (EIs) e Valente trata unicamente das colocações). A última utiliza as funções lexicais (FLs) de Mel'čuk (1998) em sua análise, mas estas funções baseiam-se em recursos como a sinonímia e antonímia e por isso são mais apropriadas para a criação de dicionários do que para o contexto da TA, como é o foco da presente pesquisa. Além disso, como as CLs também podem ser especializadas, a preocupação com a discriminação das variedades do português é também em relação às combinatórias lexicais especializadas (CLEs), que também apresentam muitas diferenças neste sentido.

Por toda a complexidade envolvida, as CLs são, indiscutivelmente, um grande desafio para a TA e também para os editores humanos, dependendo da sua língua materna. Embora muitas vezes os erros provenientes das CLs possam não afetar integralmente a compreensão do texto, eles certamente apresentam problemas de restrição combinatória que impactam na qualidade do texto e podem causar uma imprecisão que é justamente o oposto da objetividade pretendida através da comunicação por e-mail, tíquetes de suporte, entre outros conteúdos que a empresa traduz.

## **1.1 Objetivos**

**O objetivo geral deste trabalho é compreender quais problemas estão envolvidos na tradução das combinatórias lexicais do *corpus* analisado e então apresentar as possíveis soluções.**

Pretendemos alcançar este objetivo através de:

- Descrição do estado da arte para as CLs

- Verificação na plataforma de anotação nos tipos de erros “*Lexical Selection*”, “*Overly Literal*” e “*Wrong Language Variety*” e:
  - Análise do processo de anotação relativamente às CLs.
  - Análise dos erros de tradução de CLs (verificação da origem dos erros, se é de uma estrutura problemática na língua de partida (LP) ou da interferência da variedade da língua de chegada (LC))
- Sistematização de erros frequentes de CLs entre PB e PE em tarefas de pós-edição e contribuição para a detecção automática dessas variedades do português.

Quais são as características, os tipos, os problemas e as estratégias de tradução das CLs? Em quais tipos de erros os problemas que envolvem CLs são anotados atualmente? Quais são os critérios de anotação para os erros de CLs? Há erros provenientes de CLEs? Quais são as origens dos erros? Há erros de diferenças entre PB e PE nas CLs do *corpus*? Estas foram algumas das perguntas para as quais procuramos encontrar respostas.

A dissertação está organizada da seguinte forma: O capítulo seguinte apresenta a entidade de acolhimento do estágio, a Unbabel. O capítulo 3 apresenta o estado da arte sobre as CLs, explorando suas propriedades, tipos, relevância; analisando os problemas que estas unidades representam para a tradução, as implicações para a TA, etc., e apresentando as respectivas estratégias de tradução. O capítulo 4 trata da metodologia utilizada, descrevendo o processo de anotação da empresa, o *corpus* utilizado e glossário criado. O capítulo 5 trata da análise dos dados, que consiste na análise do processo de anotação e dos erros de tradução de CLs, e aponta as sugestões para a empresa. Finalmente, o capítulo 6 apresenta as conclusões.

Apresenta-se também, em anexo, um glossário que foi criado ao longo do trabalho, com as soluções encontradas para os erros de tradução do *corpus* analisado.

## 2 A entidade de acolhimento do estágio

A Unbabel é uma *startup* portuguesa que presta serviços de TA com pós-edição humana em tempo quase real. Fundada em 2013 por Vasco Pedro, João Graça, Sofia Pessanha, Bruno Silva e Hugo Silva, a empresa tem sede em São Francisco, Califórnia e escritórios em Lisboa, Nova Iorque e São Francisco.

A *startup* conta com mais de 170 funcionários que representam 27 países e 17 línguas, e recebeu os prêmios: *The Techies 2018*, na categoria *Best Use of AI for Enterprises*, o WMT16 e recentemente o WMT19 (*Best Machine Translation (MT) Quality Estimation (QE)*) e também duas vezes o prêmio de inovação TAUS (*Translation Automation User Society*), em 2015 e 2017.

A missão da empresa é construir compreensão universal. Para isso, combina TA com uma comunidade de 50.000 revisores bilíngues, atuando com 29 línguas de trabalho, com dezenas de pares de línguas entre elas. Além de manter-se flexível para abrir pares de línguas personalizados por cliente, dependendo dos volumes de trabalho em causa.

A Unbabel preocupa-se com a qualidade das traduções. A comunidade de revisores que faz a pós-edição da tradução é mantida a fim de alcançar a qualidade de uma tradução humana. Neste processo de pós-edição, os revisores são tradutores/ editores com diferentes graus de proficiência, então há várias ferramentas que os auxiliam na tomada de decisões, e após esta etapa, há outros processos, estes feitos por linguistas profissionais, que avaliam a qualidade dos revisores (processo de avaliação) e do trabalho final entregue ao cliente (processo de anotação), e há ferramentas que visam a implementação contínua de melhorias (conforme veremos mais adiante). Alcançar um padrão de qualidade excelente é primordial, de acordo com os objetivos da empresa: “*Our objective is to have all the LP's [language pairs] assessed*

(90% of the content of our pipeline) at a 95 MQM<sup>1</sup>, perceived in the industry as the professional level” (Unbabel, 2018a).

A empresa atua na área de atendimento ao cliente, traduzindo e-mails, tíquetes de suporte, *live chat* e *FAQs* (perguntas frequentes). Além disso, presta serviços de transcrição, tradução e legendagem automáticas de conteúdos de vídeo.

O tipo de sistema de TA que a empresa utiliza é *Neural Machine Translation (NMT)*. Trata-se de uma nova abordagem que veio suplantando a *Statistical Machine Translation (SMT)*. Na altura em que este trabalho foi realizado, o sistema era já NMT genérica e adaptada aos domínios e clientes. Anteriormente, a empresa utilizava o *Moses*, um tipo de *software* livre que permite treinar automaticamente modelos de tradução para qualquer par de línguas, através de textos alinhados (*corpus* paralelo). Mais detalhes sobre os sistemas de TA serão apresentados na seção 3.3.1.1.

O fluxo de trabalho da Unbabel ocorre da seguinte maneira: o cliente envia o texto de partida (TP) para a empresa, este texto passa pela TA e é dividido em diversas partes pequenas que são distribuídas entre a comunidade de revisores humanos para a pós-edição. Após esta etapa, com o auxílio de diversas ferramentas que descreveremos mais adiante, o texto pós-editado passa por um sistema de QE. Se for abaixo do limiar do sistema QE, é enviado aos editores novamente, caso seja acima, é enviado ao cliente. Há um revisor humano *Senior Editor* que revê o texto integral. A figura 1 ilustra o fluxo de trabalho da Unbabel.

---

<sup>1</sup> A métrica de qualidade utilizada pela empresa será apresentada na seção 4.1.2.

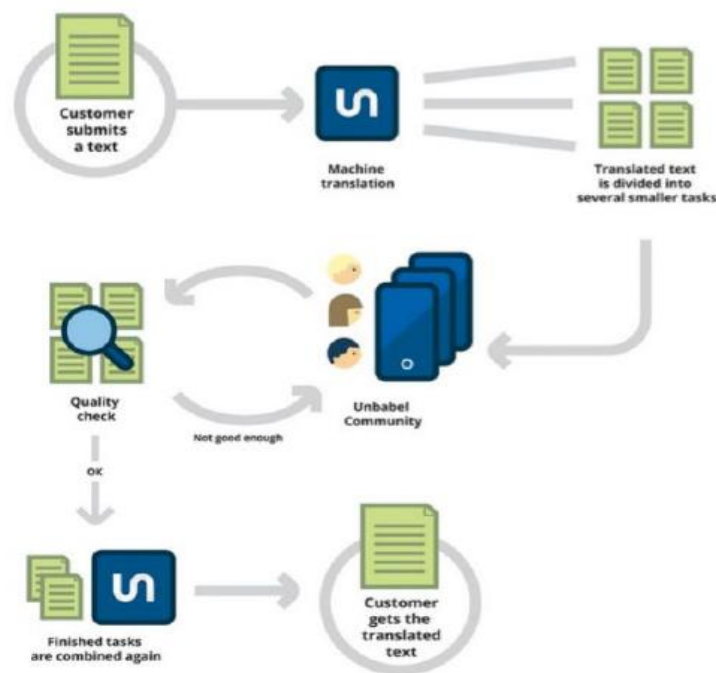


Figura 1: Unbabel's Translation Pipeline. (Extraído de Unbabel, 2018b, p.3)

Periodicamente, são realizadas anotações de erros em amostras de trabalhos realizados, através do processo de anotação da empresa, que serve para avaliar a qualidade do serviço entregue ao cliente. O presente trabalho é baseado neste último processo e a seção 4.1 é dedicada ao mesmo.

Em relação ao processo de avaliação que verifica o desempenho dos revisores, é feito pelo sistema “*Evaluate*” e funciona da seguinte maneira: os revisores são testados em três estágios, (o primeiro estágio é apenas um teste para verificar a habilidade do revisor de fazer a pós-edição, o revisor ainda não é remunerado neste estágio e não é qualificado pelo processo de avaliação). É a partir do segundo estágio que os revisores começam a ser classificados pelo processo de avaliação, quando já são remunerados. Os avaliadores são tradutores profissionais com pelo menos cinco anos de experiência e avaliam o desempenho dos revisores analisando desde o texto original e a versão extraída da TA até a última versão pós-editada; fazem um comentário sobre a pós-edição final e classificam-na dentro de uma escala de 1 a 5 (1 para o pior desempenho e 5 para o melhor).



Para assegurar a qualidade e a agilidade das traduções, existem *Guidelines* para os processos de pós-edição (específicas por língua de trabalho), de avaliação e de anotação, e os revisores, avaliadores e anotadores têm acesso, em suas plataformas, a uma série de ferramentas com informações adicionais, como indicação do tipo de registro que deve ser utilizado em cada caso (formal ou informal), instruções ou glossários de clientes, memórias de tradução (soluções para trechos que já foram traduzidos previamente) e “*Context*”, ferramenta criada pela empresa, que propicia mais algum contexto sobre o fragmento traduzido. Esta opção pode ser muito útil para evitar erros a respeito de referentes, número e gênero.

Adicionalmente, existem ferramentas para a detecção automática de erros, como o *Smartcheck*, sistema que dá dicas destacando palavras ou grupos de palavras: em vermelho quando existe um erro e em verde para sugerir melhorias no estilo, e o *Dependency Parser* (desenvolvido por Martins *et al.* (2013)), um analisador sintático que fornece informações sobre a estrutura de uma frase. De acordo com Testa (2018), o *Dependency Parser* é utilizado para analisar dados a fim de transmitir informações mais específicas para o *Smartcheck* de forma a obter correções mais precisas.

Os revisores, avaliadores e anotadores podem, ainda, utilizar a função “*Report*” em suas plataformas, para registrar problemas técnicos em qualquer destas ferramentas. A propósito, todas estas ferramentas são muito úteis para os processos da empresa fluírem mais facilmente, no entanto, por se tratar de uma *startup* e, por isso, um ambiente de constante implementação de novas ferramentas, estas são atualizadas constantemente e melhoradas com o *feedback* dos usuários.

Dado este contexto sobre a empresa, seu fluxo de trabalho, áreas de atuação, serviços e processos, a presente dissertação se concentrará no processo de anotação, como já mencionado.

O próximo capítulo apresenta o estado da arte no que diz respeito ao tipo de unidades em que nos focamos neste trabalho, as CLs.

### **3 Estado da arte**

Este capítulo apresenta o referencial teórico sobre as CLs, as características, os tipos, os problemas que suas propriedades representam para a tradução, sobretudo para a TA, e trata de algumas estratégias de tradução. O subcapítulo 3.1 apresenta as propriedades das CLs. O subcapítulo 3.2 trata da tipologia, são apresentadas quatro classificações de combinações de palavras e apontadas algumas características das colocações, das fraseologias e das EIs. O subcapítulo 3.3 trata dos problemas de tradução das CLs, descreve brevemente os principais sistemas de TA e trata dos problemas relativamente a esta área da tradução, dos problemas envolvendo as CLEs e, por último, dos impactos nas variedades linguísticas. Finalmente, o subcapítulo 3.4 apresenta algumas estratégias de tradução das CLs.

#### **3.1 Características das combinatórias lexicais**

As CLs são estruturas complexas de definir, dado que se trata de expressões ora sem qualquer ou com pouco sentido figurado (como as colocações, por exemplo: “preço exorbitante” e “café forte”), ora com significado verdadeiramente figurado (como as EI, por exemplo: “amigo da onça”) e, portanto, despertam o interesse de vários linguistas, tradutores, lexicógrafos, entre outros. “Elas apresentam-se como expressões vivas da língua, muitas vezes de cunho único, especificidades das culturas, transportadas pela inerência da sua expressividade e emotividade, expressa nos relatos de vida, cheia de “densos significados humanos, políticos e ideológicos”” (Ortiz Alvarez, 2012, p. 358).

Mas não é só a complexidade que as torna tão interessantes. Elas também estão presentes em todas as línguas: “As combinações, mais ou menos fixas, de palavras não são uma exceção, mas o caso mais normal de construção de significados nas línguas” (Ágel, 2004, p. 65 *apud* Costa, 2017, p. 17).

Outra característica destas expressões intrigantes é que não há na literatura um consenso quanto à sua designação. Segundo Ortiz Alvarez (2012, p. 358), alguns dos hiperônimos para referir as CLs são: “Fraseologia”; “unidade fraseológica”; “expressão pluriverbal”; “unidade pluriverbal lexicalizada”; “expressão fixa”; “fraseolexema”; “frasema”; “fraseologismo” e “combinatória lexical”. Alguns autores utilizam ainda a designação “combinações repetidas” para as CLs, a fim de diferenciá-las das combinatórias livres de palavras (cf. Colaço, 2014). Por outro lado, alguns autores tratam as CLs como um todo, enquanto outros estabelecem algumas distinções e consideram diferentes tipos de CLs. Mesmo neste último caso, não há unanimidade no que diz respeito à designação desses diferentes tipos de CLs.

Segundo Vilela (2002), estas unidades têm uma função nomeadora que dá força à expressividade:

Estes frasemas (ou fraseologismos) funcionam como um processo de ampliação do léxico, servindo assim para a nomeação, qualificação, circunstanciação, ou, por outras palavras, contribuindo para a lexicalização da conceptualização e categorização da nossa experiência cotidiana. (Vilela, 2002, p. 161)

O autor propõe que, mesmo nos casos em que a nomeação já existia por meio de signos primários, “o fraseologismo ou nomeia de forma mais expressiva, ou reforça a expressividade se ela já existir lexicalizada de outra forma” (idem). Assim sendo, Vilela afirma que as CLs são muito frequentes em textos predominantemente comunicativos e onde a oralidade predomina.

A lexicalização que as CLs constroem é da estrutura como um todo, uma unidade léxica completa. A combinação fixa ou semi-fixa pode ser ocasionada por recursos como a metáfora,

a metonímia, a hipérbole, etc. Além de ser característica das situações informais: “As fraseologias são uma marca da linguagem da proximidade, da oralidade, da expressividade, da descontração, da horizontalidade discursivo-pragmática” (Vilela, 2002, p. 184). Mas não é só nas situações informais que elas ocorrem. Como refere Altenberg (1998), o campo das CLs é muito vasto e, por isso, de difícil delimitação, tanto em relação ao grau de fixidez como à classificação dos diferentes tipos envolvidos. E por isso, também estão presentes nos diferentes tipos de comunicação, algumas combinatórias são comuns, inclusive, em linguagem especializada.

De acordo com Lorente *et al.* (2002), as CLEs são caracterizadas, principalmente, através de: “inclusão de um termo”, “semi-fixação” e “frequência”. Assim, Bevilacqua (2004/2005) afirma que algumas propriedades das colocações da língua comum (como veremos mais adiante) são aplicadas às unidades em textos especializados. “A diferença será que estas últimas incluem desde termos complexos ou sintagmáticos (*fractal complexo, ataque cardíaco*) até unidades maiores (*cometer um crime ambiental*)” (Bevilacqua, 2004/2005, p. 80). Assim, não é evidente a distinção entre alguns tipos de CLs e os termos complexos. A autora considera que esta é uma perspectiva ampla de fraseologia especializada, “uma vez que inclui todas unidades sintagmáticas, independentemente de algumas delas se caracterizarem como termos, como *ataque cardíaco*” (idem).

Lorente (2001) apresenta uma análise comparativa das unidades fraseológicas da língua geral e das unidades fraseológicas especializadas (UFEs). De acordo com a autora, os aspectos diferenciais estão relacionados com certas tendências, condicionadas pelo *corpus* de análise e pelas condições de comunicação em questão. A tabela 1 ilustra estas diferenças:

Tabela 1

*Diferenças entre Unidades Fraseológicas da língua geral e Unidades Fraseológicas Especializadas (Extraído de Lorente, 2001, p. 19)*

<i>UF de la LG</i>	<i>UFE</i>
Son combinaciones de palabras de la LG	Son combinaciones de términos (UT)
Las UF aportan connotación, expresividad	Tendencia mayor a la denotación
Significado general, relacionado con el conocimiento del mundo y con la experiencia.	Significado especializado, relacionado con unos conocimientos reglados y aprehendidos voluntariamente.
Aparición frecuente en el lenguaje oral, o en expresiones literarias que intentan reflejarlo.	Localización prioritaria en textos escritos, ya que la oralidad no suele ser analizada en ámbitos especializados.
Identificación con una comunidad lingüística, cultural, religiosa, o con un grupo social.	Identificación con una temática, con una comunidad académica, con una escuela o con un grupo profesional (argot).

Lorente *et al.* (2002) distinguem as CLEs em dois tipos: 1) “*la fraseologia verbal especializada (tratar la hepatitis, absorber el calor)*” (Lorente *et al.*, 2002, p. 651) e 2) “*la fraseologia nominal especializada (tratamiento de la hepatitis, absorción del calor), que es resultado de la nominalización de una unidad fraseológica verbal que, en algunos casos, llega a ser considerada unidad terminológica*” (*op.cit.*).

Segundo as autoras, há também uma tendência para a utilização de verbos de suporte (ou “verbos leves” conforme Gonçalves e Raposo (2013)) nas CLEs, que são verbos homônimos de verbos plenos/principais, mas que, quando se combinam com certos nomes (seus complementos), são afetados por um processo de esvaziamento semântico. “*En los textos especializados se puede constatar un uso abundante de nominalizaciones deverbales con verbos de soporte, en lugar de sus correlatos verbales simples (Ejemplo: realizar una saturación vs. saturar)*” (Lorente *et al.*, 2002, p. 651). Na seção 3.3.2, trataremos dos problemas de tradução das CLEs.

Por suas propriedades, as CLs diferenciam-se das combinatórias livres de palavras. Estas propriedades incluem: “(i) polilexicalidade (formadas por pelo menos dois elementos), (ii) alta frequência de ocorrência (ocorrência recorrente), (iii) institucionalização (reconhecidas

pelos falantes como naturais, familiares) e (iv) estabilidade (manutenção dos elementos que integram a unidade)” (Corpas Pastor, 1998). Como veremos a seguir, as designações para as propriedades também não são unânimes entre os autores. Passemos então à descrição das propriedades mais comuns.

### 3.1.1 Fixidez

A fixidez consiste na qualidade das CLs enquanto blocos pré-formados. Esta propriedade está interligada com os conceitos de Corpas Pastor (2001, p. 92) de “estabilidade” (manutenção dos elementos da CL) e “institucionalização” – “*Dicho rasgo debe entenderse como la fijación en función de la reproducibilidad de estas en el discurso.*” (*op.cit.*) –, que é o reconhecimento pelos falantes das expressões como naturais e familiares. E este reconhecimento é da expressão como um todo. Hotopf (1983 *apud* Corpas Pastor, 2001, p.92) aponta para a possibilidade dos erros chamados “*supressed slips*”, que são erros de reprodução linguística que consistem no emprego de uma colocação completa (*cut my nails*) no lugar da que se pretendia utilizar (*brush my teeth*). Como refere Costa (2017, p. 28), “Estas combinatórias estão disponíveis na mente como uma estrutura pré-fabricada, são psicolinguisticamente salientes, ao contrário das combinações livres, e são já conhecidas antes do ato de fala que as reproduz”. Assim, de forma geral, o que determina a fixidez de uma CL é o uso típico e recorrente da mesma.

As CLs são fixas ou semi-fixas (ao contrário das combinatórias livres de palavras, que não possuem nenhum grau de fixidez – são totalmente livres). “A fixidez é um conceito gradual em vários aspetos: no fim da escala encontramos as expressões fixas e, no outro extremo, encontram-se expressões com poucos sinais de fixidez, as colocações” (Burger, 2004, p. 21 *apud* Costa, 2017, p. 28).

A fixidez pode também ter uma dimensão pragmática, uma vez que certas CLs podem ter uma função pragmática: “[...], ou seja, fórmulas de rotina convencionadas a determinadas situações, como: Bom dia. Mas também fórmulas que regulam a comunicação, como: Estás a perceber?” (Costa, 2017, p. 16). Este tipo de CLs é muito frequente no *corpus* analisado no presente trabalho, visto que é característico dos e-mails, das saudações profissionais, etc.

### 3.1.2 Recorrência

As CLs também podem ser identificadas com base na sua ocorrência recorrente. Sinclair (1991, pp. 109–115) trata dos fatores de repetição que existem na reprodução do discurso, utilizando os termos “*open choice principle*” e “*idiom principle*”.

De acordo com Corpus Pastor (2001), os elementos integrantes de uma colocação, por exemplo, podem aparecer juntos com alguma frequência. Esta frequência pode ser apurada através da análise de *corpora* e é particularmente interessante para a TA:

*[...] computational linguists use statistical techniques applied to aligned, parallel, bilingual corpora to identify collocation translations and semi-automatically construct a bilingual collocation lexicon. Such a lexicon can then be used as part of a machine translation program. (McKeown & Radev, 2000, p. 2)*

Ou seja, esta propriedade é o fator determinante para o valor estatístico das colocações, possibilitando, assim, a utilização de programas de gestão de *corpus* e extração automática das colocações (cf. Corpus Pastor, 2001). No entanto, apesar de ser possível a extração automática das CLs, há muitas limitações nesses sistemas, conforme veremos na seção 3.3.1 que trata dos problemas de tradução envolvidos na TA.

### 3.1.3 Sentido figurado

O sentido figurado das CLs ocorre em diferentes graus, pois algumas expressões conservam o sentido literal de um ou mais dos componentes. Costa (2017) apresenta exemplos de expressões parcialmente idiomáticas: “armado até os dentes” e “tocar de ouvido”, em que um dos componentes mantém o sentido literal, e de expressões totalmente idiomáticas: “perder a cabeça”, “cantar vitória”, em que o sentido é totalmente figurado.

Apesar disso, o sentido figurado não está presente em todas as CLs. No caso das colocações, por exemplo, que são apenas semi-fixas, um dos seus constituintes pode ter sentido figurado ou não. No entanto, se todos os constituintes mantêm o sentido literal, então a escolha de um dos constituintes não é feita livremente, como veremos na seção 3.2.1.2.

Neste sub-capítulo, vimos que as CLs são complexas de definir, estão presentes em todas as línguas e que não há um consenso quanto às suas designações. Vimos ainda que algumas destas unidades têm uma função nomeadora que dá força à expressividade e que a lexicalização que estas estruturas constroem é da estrutura como um todo, uma unidade léxica completa. Também mencionamos que as CLs podem aparecer na linguagem especializada. Por fim, apresentamos algumas propriedades das CLs e mencionamos que elas se distinguem das combinatórias livres de palavras. A próxima seção apresenta a tipologia das CLs.

## 3.2 Tipologia das combinatórias lexicais

Conforme já mencionado, as designações que se encontram na literatura para estas unidades são variadas. E não é só. Os tipos de CLs também não é um aspecto unânime entre os autores. Assim, as CLs não possuem limites claros, pois as unidades lexicais abrangidas



apresentam propriedades manifestadas em maior ou menor grau, mas dificilmente uma CL contém todas as propriedades definidoras (Cowie, 1998).

Além de os autores não concordarem quanto à designação e aos tipos de CLs, também foi mencionado atrás que há ainda uma disparidade quanto a tratá-las em conjunto, associando-as a um hiperônimo, ou tratá-las como unidades individuais e estabelecer as fronteiras. Nesta pesquisa, tendo em conta os nossos objetivos, optamos por não fazer a distinção e utilizar a designação “combinatórias lexicais” para todas elas, pois, propomos para a empresa uma visão mais abrangente do fenômeno e a noção de que estas unidades não podem, via de regra, ser traduzidas literalmente e devem ser consideradas como um todo. Independentemente disso, decidimos apresentar as características de alguns tipos de CLs individualmente, pois, apesar de não haver um consenso em relação aos tipos, existem algumas diferenças relevantes entre as unidades que devem ser consideradas no processo de tradução.

Por exemplo, de acordo com Colaço (2014), Corpas Pastor distingue as colocações das fraseologias desta forma: “as colocações não constituem um enunciado (definido como uma unidade de comunicação), mas apenas uma expressão que ocorre no interior de um enunciado” (idem) e as fraseologias podem funcionar como um enunciado. Neste sentido, Colaço (2014) dá os seguintes exemplos de colocação: “dar um murro”; “assinar um contrato”; “tomar banho”, e os seguintes exemplos de fraseologias: “os meus sentimentos”; “guardo deferimento”; “agradeço desde já a atenção dispensada”.

Saber essas diferenças é fundamental para a tradução, contudo, a autora chama a atenção para o fato de que esta distinção não é o único fator determinante, pois, frequentemente, está associada à complexidade (ou mesmo extensão) da unidade e nem sempre isto caracteriza o que é colocação ou fraseologia. Mas “classificam-se como fraseologias, muitas vezes, as combinatórias com um maior grau de fixidez. Ex: ambos os textos fazendo igualmente fé salvo acordo das partes em contrário” (*op.cit*).

Baseado em grandes especialistas destas unidades, apresentamos quatro classificações de combinações de palavras. A primeira é a classificação de Hausmann (1984 *apud* Costa, 2017) conforme a figura 2. Nesta classificação, são diferenciadas, primeiramente, as estruturas fixas e as não fixas. Nas estruturas fixas, encontramos as palavras compostas e frases feitas, como “caixa-forte” e “fazer a cabeça de<sup>2</sup> alguém”. Dentro das estruturas não fixas, encontramos combinações, divididas em “*Konter-Kreation*” (*counter-creation*), “*Kollokation*” (*collocation*) e “*Ko-Kreation*” (*co-creation*). As “*co-creation*” são as combinatórias de palavras livres e as “*collocation*”, as colocações “são produtos “semi-prontos” da língua, o falante não as pode alterar criativamente, vai buscá-las à memória como um todo e o ouvinte reconhece-as” (Costa, 2017, p.19). Por fim, as “*counter-creation*” são combinações nas quais um elemento é selecionado de acordo com o seu significado, mas o número de objetos extralinguísticos é restrito e por isso a capacidade combinatória é limitada por motivos semânticos (diferente das colocações que, como veremos mais detalhadamente na seção 3.2.1.2, as possíveis combinações são determinadas essencialmente através do uso e não de uma motivação semântica). “Exemplos deste fenómeno seriam combinações do género de um rebanho de ovelhas ou um nariz aquilino” (cf. Hausmann, 1984, pp. 398-399 *apud* Costa, 2017, p. 18).

---

<sup>2</sup> “Fazer a cabeça a alguém” no exemplo de Costa (2017), em PE.

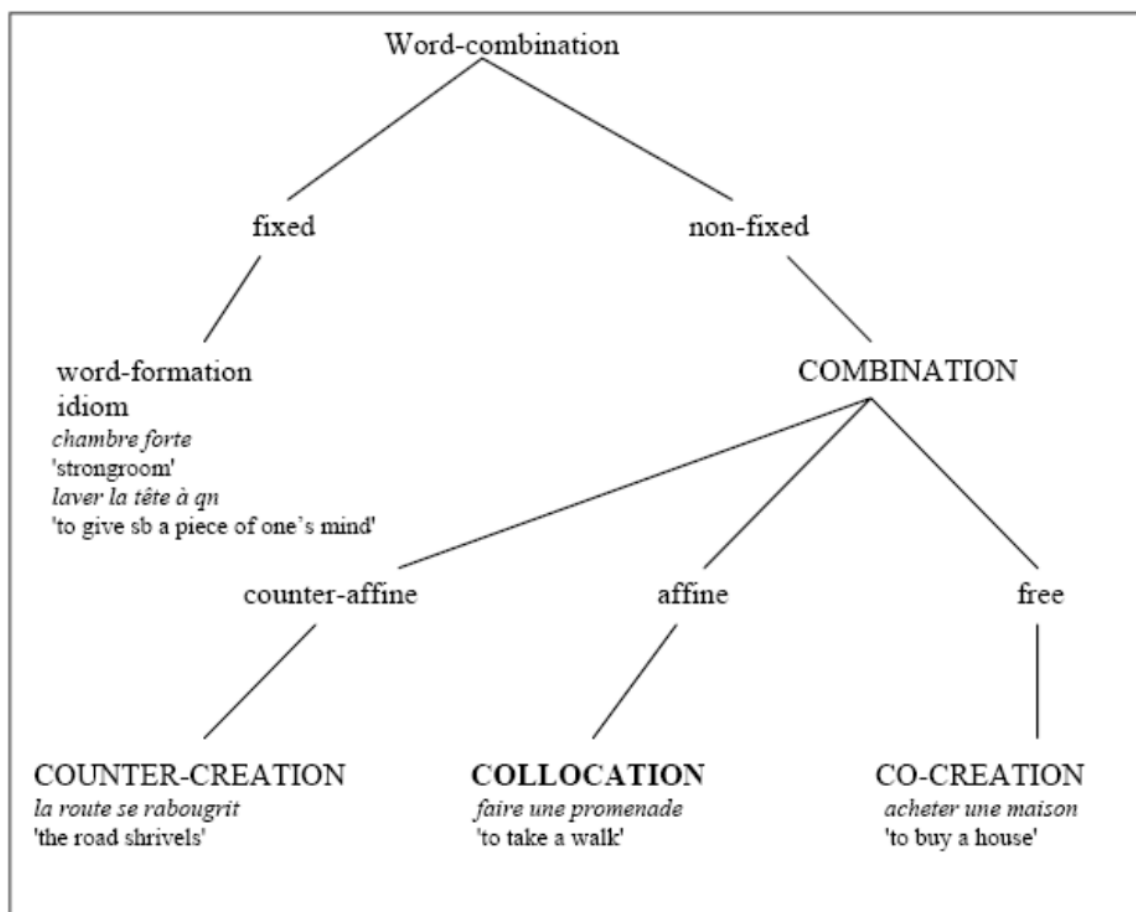


Figura 2: Classificação de combinações de palavras de Hausmann (1984 *apud* Costa 2017, p. 19).

A segunda classificação apresentada é a de Mel'čuk (1998), representada na figura 3. Este autor faz uma divisão entre frasemas semânticos e frasemas pragmáticos. Os frasemas semânticos são divididos em frasemas completos (EIs), semi-frasemas (colocações) e quase-frasemas (fraseologias).

Nos frasemas completos, na expressão “bater as botas”, por exemplo, o significado não resulta da soma dos dois elementos “bater” (A) e “botas” (B), em vez disso, há um sentido diferente (C) “morrer” que não inclui A, nem B. Nos quase-frasemas, é mantido o significado individual dos lexemas que os compõem, mas é acrescentado um novo sentido que não é dedutível da soma das partes constituintes (cf. Sanromán, 2000, p. 167). Por exemplo, em PE, “cinturão negro [ou faixa preta, no PB]: que além de “cinto” e “negro”, acrescenta-se um

sentido: “determinado grau de conhecimento e habilidades marciais”. Por fim, os semi-frasemas aqui são as colocações.

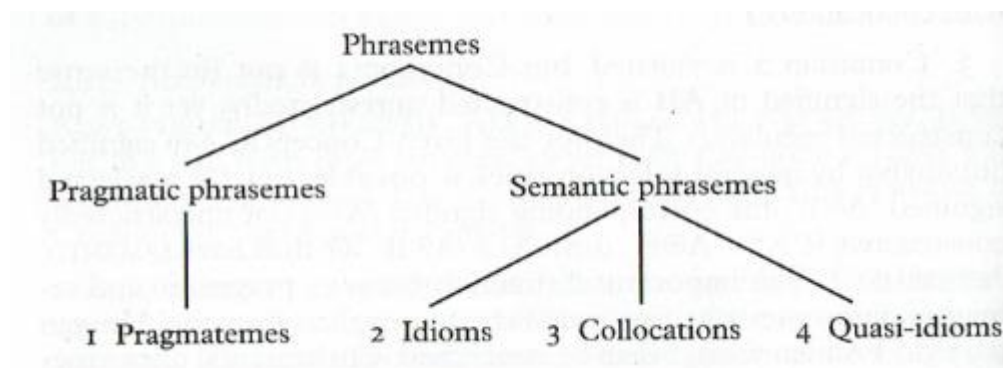


Figura 3: Classificação de combinações de palavras de Mel'čuk (1998)

A terceira classificação é a de Cowie (2001[1998] *apud* Costa, 2017), apresentada na figura 4. O autor divide as combinações de palavras em dois grandes grupos: *composites* e *formulae*. *Formulae* são combinações com função essencialmente pragmática, como “Bom fim-de-semana” ou “Muitos parabéns” (cf. Costa, 2017). No grupo *composites*, as estruturas são distinguidas com base em dois critérios: o da transparência (sentido literal ou não) e o da comutabilidade (grau de comutação dos elementos da combinatória). Aqui as combinatórias são divididas em três grupos: (i) colocações restritas (as colocações), (ii) idiomas figurativos (fraseologias) e (iii) idiomas puros (as EIs).

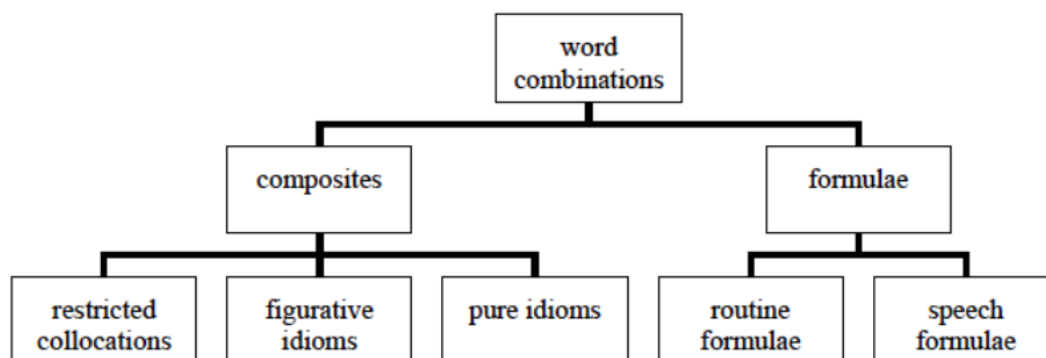


Figura 4: Classificação de combinações de palavras de Cowie (2001[1998] *apud* Costa, 2017, p. 22)

A quarta e última classificação é a tipologia de Burger (2003 *apud* Costa 2017) – veja-se a figura 5 - e foi a adotada por Costa (2017). Nesta classificação, a divisão inicial leva em consideração a função que as CLs desempenham na comunicação. Desta forma, é estabelecida uma distinção entre a) fraseologias de valor referencial, b) fraseologias estruturais e c) fraseologias de valor comunicativo.

As combinatórias referidas em (a) são capazes de “denotar objetos ou “estados de coisas” da realidade extralinguística. Como exemplo, podemos aduzir expressões como *Em casa de ferreiro, espeto de pau, deitar achas para a fogueira, sorriso amarelo* (provérbios, idiomatismos ou colocações, respectivamente)” (Burger, 2003 *apud* Costa, 2017, p. 22). As combinatórias do tipo (b) assumem uma função essencialmente gramatical (as locuções), por exemplo, “em relação a” e “assim como”. Por último, (c) envolve as expressões que têm uma função pragmática como “Tanto quanto sei...”, “Com os meus melhores cumprimentos...”, “Era uma vez...”, “Feliz Natal!”.

As “fraseologias de valor referencial” são ainda subdivididas por meio de um critério semântico. Essa subdivisão é entre: fraseologias proposicionais (os provérbios) “com capacidade para realizar asserções sobre objetos e “estados de coisas” (“Em casa de ferreiro

espeto de pau”, “Até o diabo se ria”)” (*op.cit.*) “a galinha do vizinho é sempre melhor do que a minha” e fraseologias denominativas, que referem/denominam objetos e “estados de coisas”. Por exemplo, um “sorriso amarelo” é a denominação para um sorriso pouco natural (cf. Costa, 2013).

Ainda de acordo com Burger (2013 *apud* Costa, 2017), as fraseologias proposicionais são subdivididas através de critérios sintáticos e textuais entre frases fixas (que constituem réplicas e, portanto, menor autonomia, como “Só faltava esta agora!”), e fórmulas tópicas (enunciados genéricos, como os provérbios (“Cachorro que late não morde”) e os lugares-comuns (“Ninguém é perfeito”, “A vida é curta”), com maior independência.

E quanto às fraseologias denominativas, que estão mais próximas das combinatórias abordadas neste trabalho, o autor faz uma subdivisão semântica, conforme o grau de idiomaticidade: as fraseologias idiomáticas como “ficar fora de si”, as parcialmente idiomáticas, como “jurar de pés juntos” (PB) ou “jurar a pés juntos” (PE) e as não ou pouco idiomáticas como “escovar os dentes”, estas últimas o autor inclui no conceito genérico de colocação.

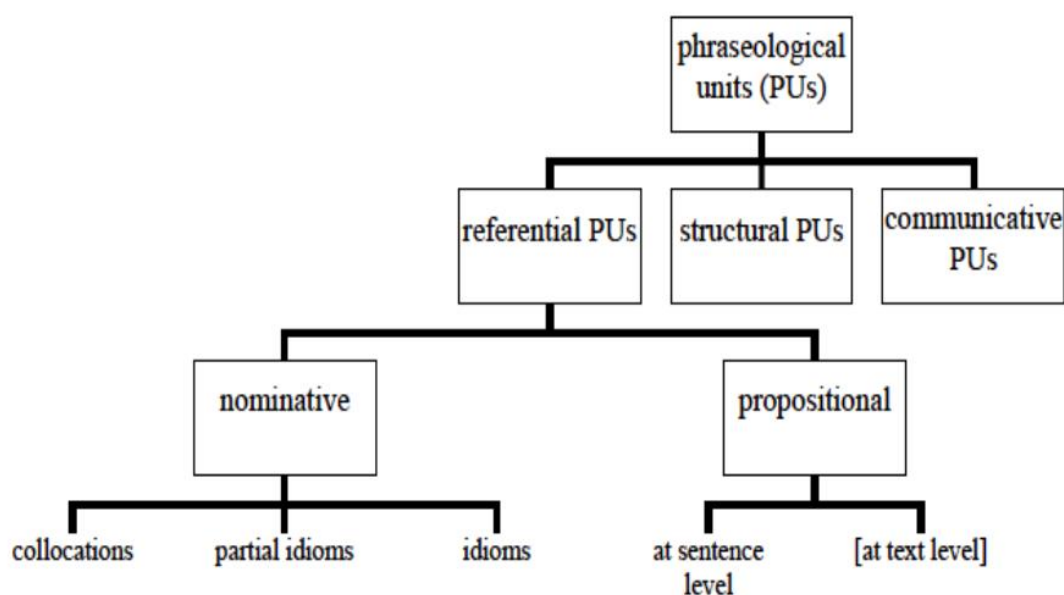


Figura 5: Classificação de combinações de palavras de Burger (2003 *apud* Costa, 2017, p. 23).

Assim, de uma forma geral, muitos autores concordam que as CLs diferem quanto ao grau de fixidez. E também quanto à sua função.

Cowie (1981) nos remete à escala gradual de CLs para distingui-las, ao utilizar os critérios de idiomatismo e restrição combinatória. O autor concorda que a colocação é a unidade menos fixa na escala gradual das CLs “*A collocation is by definition a composite unit which permits the substitutability of items for at least one of its constituent elements (the sense of the other element, or elements, remaining constant)*” (Cowie, 1981, p. 224). No exemplo do autor, “*run a business*”, o colocativo “*run*” adquire o significado de “*be in charge of*” quando se coloca com bases que denotam negócios ou qualquer atividade comercial, como teatro, restaurante, etc, possibilitando, assim, as colocações “*run a theatre*”, “*run a restaurant*”, etc, desde que a substituição fique restrita a esta classe semântica de substantivos.

Para os investigadores McKeown & Radev (2000, p. 3), contudo, o exemplo é considerado como combinatória livre de palavras:

In the case of *run* (in the sense of ‘manage’ or ‘direct’) the semantic restrictions on the object are tighter but still follow a semantic generality: any institution or organization can be managed (e.g. *business, ice cream parlor*, etc.). **In contrast to these free word combinations**, a phrase such as *explode a myth* is a collocation. In its figurative sense, *explode* illustrates a much more restricted collocational range. (McKeown & Radev, 2000, p. 3) [negrito meu]

Isso mostra como estas expressões são complexas e difíceis de definir e classificar ou até mesmo de identificar.

Todavia, Mckeown & Radev (2000) também concordam que a fixidez das CLs é gradual, e fazem uma distinção entre as colocações e as combinatórias livres de palavras. Para os autores, as colocações estão situadas entre as combinações livres de palavras e as EIs. Sendo as últimas definidas como expressões cujo significado não pode ser dedutível do significado dos seus constituintes individualmente, ao contrário das colocações, em que o significado dos constituintes pode contribuir para o significado da unidade; e as combinações livres de palavras são definidas como aquelas em que as palavras podem ser substituídas por outras da mesma classe, desde que exista compatibilidade semântica. Mais exemplos de Mckeown & Radev (2000) em inglês são apresentados na tabela 2 a seguir.

Tabela 2

*Collocations vs. idioms and free word combinations. (Extraído de McKeown & Radev, 2000, p. 3)*

Idioms	Collocations	Free word combinations
to kick the bucket	to trade actively	to take the bus
dead end	table of contents	the end of the road
to catch up	orthogonal projection	to buy a house

Por tudo isso, o estudo das CLs não é uma tarefa simples.

Nesta seção, referimos que os tipos de CLs não são uniformes entre vários autores, assim como a opção de tratá-las em grupo ou individualmente, e apresentamos quatro propostas de classificações de combinações de palavras. Independente destas divergências no tratamento das CLs, vimos que a fixidez é um fator considerado determinante na distinção destas expressões.

Passemos à apresentação individual dos tipos de CLs mais relevantes para o presente trabalho.



### 3.2.1 Colocações

As colocações são definidas principalmente através do uso. São expressões que dependem de contexto de comunicação e nem sempre podem ser explicadas através de regras semânticas ou sintáticas. Por este motivo, para a TA dessas expressões, é fundamental uma intervenção humana, pois as propriedades de caráter meramente estatístico nem sempre são suficientes para identificá-las e traduzi-las adequadamente.

Alguns autores definem “colocação” essencialmente através dos critérios estatísticos (Sinclair (1991), Mckeown & Radev (2000)). A primeira definição desta linha foi dada por Firth (1968, p. 99 *apud* Corpus Pastor, 2001, p. 97). Porém, o autor classifica a colocação como uma mera combinação de palavras “*Collocation states the habitual company a key-word keeps*” (*op.cit.*).

Entretanto, alguns anos depois surgiu uma concepção mais restrita do fenômeno “colocação”. Assim, Costa (2017) utiliza a mesma definição que Grossmann e Tutin (2002) “a co-ocorrência lexical privilegiada de dois (ou mais) elementos linguísticos que entre si estabelecem uma relação sintáctica” (Grossmann & Tutin, 2002, p. 3 *apud* Costa, 2017, p. 26). Ou seja, os componentes da colocação não são determinados apenas pela sua recorrência (como pode ser o caso das combinatórias livres de palavras). Mas, existe uma restrição combinatória entre os elementos de uma colocação. E a investigadora defende que a preferência por determinado item lexical, ao invés de um sinônimo, é dada através do uso e não a nível sintático ou ao nível do significado conceitual. Ao contrário da co-ocorrência lexical livre, que é uma questão semântica, a co-ocorrência lexical restrita, fenômeno em que se enquadram as colocações, é uma questão meramente lexical, determinada pelo uso (cf. Sanromán, 2000, p. 170).

Por causa da estrutura sintática, pode haver uma distância grande entre os elementos da colocação. Não há na literatura um consenso quanto a essa distância colocacional. Corpus

Pastor (2001, p. 101) menciona um exemplo de Howarth (1996, p. 71), tirado do jornal *The Guardian*, em que os colocados *have-impact* aparecem separados um do outro por 28 palavras. Conforme veremos posteriormente, isto representa um grande problema no âmbito da TA.

Em relação às propriedades das colocações, Corpas Pastor (2001) afirma que as colocações devem, por direito, ser tratadas no campo das CLs, pois possuem propriedades em comum. A autora mostra a natureza fraseológica das colocações (como a característica polilexicalidade, por exemplo) evidenciando assim a necessidade de as incluir nos estudos das CLs e apela para a importância desta inclusão. A mesma dá os seguintes exemplos de colocações, em espanhol: “*soltero empedernido*” e “*lluvia torrencial*”. Há, contudo, uma propriedade que é, segundo alguns autores, exclusiva das colocações, a transparência semântica, conforme veremos a seguir.

### **3.2.1.1 Transparência semântica**

A transparência semântica, que é o sentido não figurado das expressões, é um aspecto considerado específico das colocações (cf. Grossmann & Tutin, 2002 *apud* Costa, 2017, p. 28). No exemplo de Costa (2017), a expressão “bater a bota” não é interpretada no sentido literal, trata-se de uma expressão fixa (provérbios, frases feitas, etc.), enquanto uma expressão como “leite azedo” é uma colocação, pois trata-se mesmo do leite que está azedo.

Corpas Pastor (2001, p. 103) refere os conceitos de “verdade”, “tipicidade” e “precisão semântica” apontados por Koike (2000), de acordo com os quais as colocações devem ser típicas e verdadeiras, de forma que os colocados sejam interpretáveis no mundo real. A autora dá o exemplo, em espanhol, “*cargar una pistola*”, que é uma colocação. Ao contrário de “*lavar*” ou “*olvidar*” *una pistola*, pois o substantivo “*pistola*” só poderia estabelecer uma relação típica na qualidade de “arma de fogo”.

Mas, apesar de a transparência semântica, é importante salientar que as colocações podem ter um sentido metafórico:

*Precisamente, muchos dobletes fraseológicos se deben a la especialización de carácter metafórico (o metonímico). Nos referimos a UFS que constituyen colocaciones y locuciones a la vez, como meter un gol, tragar saliva o abrir una brecha, debido "a la metaforización de sus vínculos colocacionales correspondientes". (Koike, 2000, p. 32 apud Corpas Pastor, 2001, p. 93)*

Tais problemas serão tratados mais detalhadamente no subcapítulo 3.3.

Logo, os conceitos que a autora menciona (verdade, tipicidade e precisão semântica), apontam no sentido de que as colocações são de fácil reconhecimento para os falantes. No seu exemplo, em vez de se falar “tentar lembrar alguém de alguma coisa esquecida”, utiliza-se a colocação “*refrescarle la memoria*”, em espanhol.

Estas unidades apresentam ainda algumas restrições especiais, conforme mostra a seção a seguir.

### **3.2.1.2 O papel da restrição combinatória**

Já foi referido atrás que as CLs possuem a propriedade “estabilidade”. No caso das colocações, essa estabilidade tem a ver com os termos “arbitrariedade” e “restrição combinatória”. Consiste na impossibilidade de substituir algum componente das colocações por um sinônimo. “*The notion of **arbitrariness** captures the fact that substituting a synonym for one of the words in a collocational word pair may result in an infelicitous lexical combination*” (McKeown & Radev, 2000, p. 3). No exemplo dos autores, em inglês, “*make an effort*” é uma colocação, mas “*make an exertion*” não.

Colaço (2014) dá o exemplo “beber leite” e “tomar café” (em vez de “tomar leite” e “beber café” respectivamente), em que apenas o significado das palavras não garante as combinações mais frequentes. Da mesma forma que não é usual dizer que “o leite estava podre” (e não azedo) e que “o ovo estava rançoso ou amargo” (e não podre), por exemplo.

Assim como no exemplo de Hausmann (*apud* Costa 2017), não existe uma motivação semântica que possa explicar o fato de “torrencial” aparecer em co-ocorrência com “chuva”. O autor afirma que o fato de as colocações incluírem a transparência semântica não as impede de serem imprevisíveis.

Sanromán (2000) dá o seguinte exemplo: se quisermos exprimir o sentido de “ódio intenso”, a escolha do adjetivo para a palavra “ódio” não fica a nosso critério, pois o adjetivo “mortal” é imposto para esta palavra.

Através deste critério, (Cowie, 2001[1998] *apud* Costa, 2017, p. 33) diferencia as colocações das combinatórias livres de palavras: “[...] as combinações nas quais uma possível restrição na substituição dos elementos é devida a propriedades semânticas, estas são chamadas combinatórias livres [...] e as combinações cujas restrições são, até um certo ponto, arbitrárias – as colocações” (*op. cit.*).

Assim, as colocações, apesar de serem recorrentes em contexto, não incluem itens lexicais que ocorrem frequentemente, (como os exemplos de Mckeown & Radev (2000), em inglês, “go” e “know” - essas são combinatórias livres que se encaixam com qualquer objeto). No caso das colocações, existe uma afinidade entre as palavras: “[...] uma expressão do género de *café fraco*, é uma colocação, pois existe restrição na escolha do adjetivo *fraco*, uma vez que não se pode dizer *café débil*, apesar de *débil* ser um sinónimo de *fraco*” (Costa, 2017, p. 28).

Entretanto, conforme mencionamos que a fixidez das CLs ocorre em diferentes graus, na colocação normalmente um dos elementos é mais restrito do que o outro e por isso não pode ser substituído. Há casos, então, em que as colocações possibilitam variações (cf. Corpas

Pastor, 2001). Como no referido exemplo, em inglês, “*run a business*”, no qual “*run*”, o colocativo<sup>3</sup>, adquire o significado, a função léxica, de “*be in charge of*” quando se coloca com bases que denotam negócios ou qualquer atividade comercial, como teatro, restaurante, empresa de transporte, etc. Sendo, por isso, possível as colocações “*run a theatre*”, “*run a restaurant*”, etc.

De acordo com Aisenstadt (1979 *apud* Fontenelle, 1994, p. 42), são as variações das colocações que as diferenciam das expressões fixas. Já que, para o autor, contrariamente às últimas, as colocações não formam uma unidade semântica e possuem um certo grau de variabilidade.

Apesar de algumas colocações permitirem esta variação de elementos, há também uma restrição quanto à escolha dos elementos nestes casos. No exemplo de Cowie acima, “*run a business*”, “*run a restaurant*”, a substituição fica limitada a esta classe semântica de substantivos. Os elementos que podem ser substituídos devem respeitar o contexto colocacional, deve haver uma relação de especialização semântica, de forma que alguns colocativos podem ser colocados com mais sinônimos do que outros.

Isto permite que as colocações possam ganhar sentidos metafóricos, por exemplo: “*Otro ejemplo de especialización metafórica lo hallamos en colocaciones como atacar posiciones, desarmar una teoría, palabras hirientes, lucha dialéctica, razonamiento débil, cuyo dominio fuente sería "Discutir es luchar"*” (Corpas Pastor, 2001, p. 93). Pois as colocações podem ser especializadas e podem também ser mais ou menos idiomáticas, dependendo das possibilidades de produzir colocações por substituição por sinônimos, etc.

As colocações também são arbitrárias/ restritas em relação às diferentes línguas, variedades linguísticas e dialetos. Colaço (2014) dá o exemplo “dar um passeio”, em português,

---

<sup>3</sup> Os conceitos de “base” e “colocativo” foram definidos por Hausmann (1985), nos quais a “base” é a palavra que determina o outro elemento co-ocorrente e o “colocativo” corresponde ao elemento determinado. Mais informações sobre estes conceitos são exploradas adiante.

que em francês equivale a “*faire une promenade*” e em inglês a “*take a walk*”. Considerando o PB e o PE, há várias diferenças neste sentido, como veremos mais adiante. Tal é o caso da expressão “jogar fora” no PB, que no PE se diz “deitar fora”. Isto causa problemas de tradução e os problemas envolvendo as variedades linguísticas, especificamente, serão explorados na seção 3.3.3.

Assim, de acordo com Corpas Pastor (2001), as colocações contêm propriedades complexas, polivalentes, que ocasionaram abordagens opostas, mas complementares.

Numa tentativa de sistematizar todos esses vínculos semânticos, surgiram as FLs da Teoria Sentido-Texto (TST) de Mel’čuk e seus colaboradores (cf. Corpas Pastor, 2001, p. 103). Esta teoria contribuiu para a criação de dicionários de colocações que, assim como as ferramentas computacionais, apresentam muitas limitações e ainda precisam de muita investigação e melhorias nesse sentido. Pois, apesar de as FLs buscarem capturar as propriedades das colocações, algumas focam apenas relações semântico-lexicais básicas, tais como a sinonímia, a antonímia e a hiponímia. Desta forma, as FLs incluem todas as relações sintagmáticas (vertical) e paradigmáticas (horizontal) da palavra entrada no dicionário. Contudo, este tipo de descrição tornaria cada entrada no dicionário muito extensa (Fontenelle, 1994, p. 46).

Na próxima seção, introduzem-se mais informações sobre as formas de sistematização das colocações.

### **3.2.1.3 Tipos de colocações**

De acordo com Costa (2017), a classificação geral dos principais tipos de colocações foi apontada por Hausmann (1985). Este autor introduziu os conceitos de “base” e “colocativo”: “Hausmann, o responsável por esta observação, chama base (“Basis”) à palavra

que determina a escolha do outro elemento co-ocorrente [...] e colocativo (“Kollokator”) ao constituinte determinado” (Costa, 2017, p. 36).

Com a teoria de Hausmann (1985, p. 119 *apud* Costa, 2017, p. 36-37), a hierarquia base / colocativo fica muito clara no processo de produção linguística de um falante. Costa (2017) dá o exemplo desta teoria, adaptado ao PE: alguém que queira produzir um enunciado com o significado “fumador que fuma muito” não tem dificuldade em usar a palavra “fumador”, mas sim em encontrar o adjetivo “inveterado”, se não for um falante nativo. Assim, os maiores problemas de produção das colocações estão na identificação do colocativo.

McKeown & Radev (2000, p. 6) confirmam que a base carrega mais do significado da colocação e determina o uso do colocativo. *“This distinction is best illustrated by collocations which include “support” verbs: in the collocation take a bath, bath is the base and the support verb take, a semantically empty word in this context, the collocator”* (McKeown & Radev, 2000, p. 6).

Poulsen (2005) apresenta alguns exemplos de colocações, em inglês, na tabela 3, em que as palavras “base” estão destacadas em negrito.

Tabela 3  
*Syntactic patterns of collocation (Extraído de Poulsen, 2005, p. 14)*

1. <b>NOUN + VERB</b>	<i>the <b>water</b> runs, the <b>candle</b> burns, the <b>gap</b> widens</i>
2. <b>VERB + NOUN</b>	<i>make a <b>deal</b>, run a <b>deficit</b>, drill a <b>hole</b></i>
3. <b>ADJECTIVE + NOUN</b>	<i>a rapid <b>increase</b>, a deep <b>cut</b></i>
4. <b>COUNT NOUN + OF + UNCOUNTABLE NOUN</b>	<i>a stroke of <b>luck</b>, a peal of <b>thunder</b></i>
5. <b>PREMODIFYING NOUN + NOUN</b>	<i>a trade <b>gap</b></i>
6. <b>PREMODIFYING PARTICIPLE + NOUN</b>	<i>a negotiated <b>settlement</b>, a fishing <b>rod</b></i>
7. <b>ADVERB + ADJECTIVE</b>	<i>deeply <b>unhappy</b></i>
8. <b>ADVERB + VERB</b>	<i>to <b>wound</b> sb deeply, to be deeply <b>wounded</b></i>

Desta forma, em relação à estrutura sintática, existem algumas regularidades:

Apesar de [...] as colocações se caracterizarem por um certo grau de arbitrariedade na associação dos seus constituintes, é relevante dizer que, do ponto de vista formal, e devido à sua parcial fixidez, as colocações seguem determinados padrões sintáticos regulares. Por exemplo, as colocações construídas em torno de um Nome incluem Adjetivos e Verbos, mas provavelmente não Advérbios e Conjunções. (Grossmann & Tutin, 2002, p. 3 *apud* Costa, 2017, p. 37)

Segundo McKeown & Radev (2000), há distinções entre colocações gramaticais e colocações semânticas: “***Grammatical collocations*** often contain prepositions, including paired syntactic categories such as verb+preposition (e.g. come to, put on), adjective+preposition (e.g. afraid that, fond of), and noun+preposition (e.g. by accident, witness to)” (McKeown & Radev, 2000, p. 5). Nestes exemplos, as palavras de classes morfológicas são as bases, que determinam as palavras com que podem ser colocadas, os colocativos. “***Semantic collocations*** are lexically restricted word pairs, where only a subset of the synonyms of the collocater can be used in the same lexical context.” (McKeown & Radev, 2000, p. 6). Exemplos desse tipo de colocações podem ser os já mencionados casos que podem ocorrer em linguagem especializada, como “run a restaurant”, “run a theater”, nos quais a substituição fica limitada a esta classe semântica de substantivos (*business*).

Esta seção apresentou as colocações, que são definidas essencialmente através do uso e não de apenas critérios estatísticos, podem envolver uma distância entre os seus elementos, possuem a propriedade “transparência semântica” (sentido não figurado) e, apesar disso, apresentam a restrição combinatória, ao contrário das combinações livres de palavras. Também foi sublinhado que conforme a fixidez das CLs ocorre em diferentes graus, as colocações



permitem variações de um dos elementos, porém, a escolha deste elemento não é feita livremente. Tais variações possibilitam que as colocações tenham sentido metafórico e sejam arbitrárias também em relação às variedades linguísticas. Foram também referidas as FLs da TST de Mel'čuk e suas limitações. Por fim, foi apresentada a classificação das colocações de acordo com Hausmann (1985), com os conceitos de “base” e “colocativo” em que a maior dificuldade recai sobre a seleção do colocativo, e foi introduzida a noção de colocações gramaticais e colocações semânticas.

Passemos para as Fraseologias.

### **3.2.2 Fraseologias**

Conforme mencionado anteriormente, as fraseologias, na qualidade de CLs, apresentam as seguintes propriedades: polilexicalidade, fixidez, recorrência e estabilidade. No entanto, alguns autores defendem que, ao contrário das colocações, as fraseologias não podem ser determinadas pela transparência semântica, já que, além do significado individual dos componentes, a expressão ganha um novo significado independente. Como no já referido exemplo: “cinturão negro” em PE [faixa preta, no PB], em que além do significado de cinto (A) e de negro (B), há um terceiro sentido (C) “grau de conhecimento ou habilidade nas artes marciais”. Vilela (2002, p. 178) dá também o exemplo: “tecto falso” que, além de “tecto” e “falso”, inclui o sentido de “para isolar acústica e termicamente”.

Em relação à fixidez, as fraseologias são consideradas muito próximas das EI. “Como no caso das expressões idiomáticas, e ao contrário das colocações ou semi-frasemas, a não produtividade dos quase-frasemas leva-nos a considerá-los como unidades lexicais e não como combinação de várias unidades” (Sanromán, 2000, p. 167).

No entanto, diferentemente das EIs, as fraseologias podem ser de linguagem comum ou especializada, bem como as colocações podem ser.

Por fim, conforme já referido, outros autores classificam como fraseologias expressões que podem funcionar como um enunciado. Como nos exemplos: os meus sentimentos; aguardo deferimento; agradeço desde já a atenção dispensada (cf. Corpas Pastor, 1998 *apud* Colaço 2014).

Nesta seção, apresentamos as fraseologias, referimos que, relativamente às propriedades, alguns autores consideram que as fraseologias não possuem a transparência semântica como as colocações, e quanto à fixidez, estão mais próximas das EIs do que das colocações. No entanto, assim como as colocações, as fraseologias podem ser de linguagem comum ou especializada. Ressaltamos que estas expressões podem ter uma função pragmática e alguns autores consideram como fraseologias unidades que podem funcionar como um enunciado.

A próxima seção trata das EIs, último tipo de CL que será apresentado neste trabalho.

### 3.2.3 Expressões Idiomáticas

De acordo com Vilela (2002), a idiomaticidade é o traço prototípico das EIs, mas a sua definição está interligada com a tradução das mesmas: “[...] a definição mais comum de **expressão idiomática** ou **idiomatismo** é dada como a sequência que não pode ser traduzida literalmente para outra língua, isto é, não é possível a tradução palavra por palavra [...]” (Vilela, 2002, p. 176).

Assim, as EIs são as CLs mais fixas entre as combinatórias apresentadas. O autor afirma: “Sabemos que nas expressões congeladas não há possibilidade de substituição por sinónimos, como *estar à mão de semear* e *\*estar à mão de colher*” (Vilela, 2002, p. 168). São totalmente idiomáticas, por exemplo, “perder a cabeça”, “bater no fundo” ou “cantar vitória”.

Nestes casos, os significados não são obtidos através da interpretação dos constituintes, visto que são significados figurados.

Trata-se pois de expressões coesas, que assumem pouca variação formal e que, no plano semântico, se distinguem claramente das colocações, uma vez que, tipicamente, na construção das colocações, pelo menos um dos elementos mantém o seu sentido habitual. (Costa, 2017, p. 20)

São muitas as finalidades das EIs. Por exemplo, através do seu uso, “O falante procura recursos para tornar sua fala mais persuasiva, como em: Não “feche os olhos” para os problemas ambientais. Ou para ficar mais próximo de seu interlocutor: [...] estamos todos “no mesmo barco”” (Camacho, 2008, p. 17). A este propósito, Jorge (2001) afirma: “As expressões [idiomáticas] integram o melhor sistema de símbolos para representar uma cultura. As expressões constituem um objeto importante da língua, uma manifestação de um saber plural (...), facilitam a comunicação, estabelecem com os outros falantes da língua uma certa convivência, uma partilha.”

Diferente das colocações e das fraseologias, as EIs não são tão comuns em linguagens especializadas (cf. Colaço, 2014), embora dependa do tipo e gênero textuais. E assim como em algumas fraseologias, não existe a transparência semântica característica das colocações: “[...] a interpretação do significado de uma expressão idiomática não se pode só basear numa leitura literal dos elementos que a compõem, existe sempre uma leitura fraseológica” (Costa, 2017, p. 16).

De acordo com Sinclair (1991), a principal diferença entre as EIs e outras CLs (como as colocações ou as fraseologias) é que as EIs funcionam como uma unidade em termos de sentido, não podem ser interpretadas com base no significado dos elementos que as compõem.

Deste modo, o significado de uma EI (muitas vezes metafórico) é obtido globalmente e não composicionalmente. Apontando também neste sentido, Jorge (2002) afirma: “A tradução das expressões idiomáticas, e das lexicalizações em geral, levanta um sem número de problemas ao tradutor. Por um lado, o leitor-tradutor tem de identificar e reconhecer as estruturas lexicalizadas na língua de partida. Por outro lado, tem de transpor essa lexicalização para a língua de chegada, tentando preservar na outra língua os mesmos efeitos do texto original”.

Nesta seção, apresentamos as EIs, caracterizadas pelo alto grau de idiomaticidade e definidas como unidades que não podem ser traduzidas literalmente. São utilizadas para tornar a fala mais persuasiva ou ficar mais próximo do interlocutor, por exemplo. Em comparação com as colocações e fraseologias, vimos que as EIs são as expressões mais fixas, não são comuns em linguagem especializada e não possuem a transparência semântica como as colocações.

É preciso observar que são várias as razões pelas quais as CLs são um desafio para a tradução. As colocações, as fraseologias e as EIs foram apresentadas individualmente, pois existem algumas diferenças que devem ser conhecidas no processo de tradução. As EIs, por exemplo, não são comuns em linguagem especializada; as colocações permitem uma variabilidade (ainda que restrita à uma mesma classe semântica de substantivos) que já na EI não é possível e ao contrário dos outros dois tipos, contêm a propriedade “transparência semântica”; e as fraseologias, por vezes, são confundidas com as EIs, de forma que não é fácil estabelecer as fronteiras.

A partir deste ponto, entretanto, a distinção entre diferentes tipos de CLs não será feita sistematicamente, pois não é essencial para a presente pesquisa e sim a noção de todos os impactos que elas podem causar para a tradução. Como afirma Sanromán (2000),

[...] o que interessa, mais do que as classificações que a lexicologia possa fazer de determinadas combinações lexicais, é como fazer a inventariação, o tratamento e a recuperação de toda a informação relativa às combinações lexicais que não possam ser traduzidas palavra por palavra, de tal modo que o utilizador saiba como utilizá-las no discurso. (Sanromán, 2000, p. 174)

Neste sentido, é preciso lembrar que são expressões que devem ser consideradas como um todo e, portanto, traduzi-las para equivalentes que também são expressões consideradas como um todo. Afinal, mais fixas ou menos fixas, a tradução literal nem sempre é a mais ideal.

Por outro lado, reconhecer e produzir as CLs são grandes desafios para a tradução. As CLs são unidades complexas de definir e, muitas vezes, até mesmo de identificar. As suas propriedades causam o maior problema neste sentido, devido ao fato de que na tradução há, normalmente, uma língua não nativa. Mesmo no caso das colocações que incluem a transparência semântica e possibilitam uma variabilidade nos seus elementos que não é possível nas expressões mais fixas, elas são arbitrárias e podem ganhar sentidos metafóricos que podem não ser os mesmos na outra língua. No exemplo, em inglês, “*run a business*”, “*run*” obteve o sentido de “*manage*”, o que não acontece em português com os verbos correspondentes. Por isso, apesar de menos fixas, as colocações também são expressões complexas e desafiadoras para a tradução. O que implica que conhecimento de uma CL em uma língua não permite, diretamente, a sua produção em outra língua.

Com este contexto, o próximo subcapítulo apresenta, mais detalhadamente, os problemas de tradução que envolvem as CLs.

### 3.3 Problemas de tradução das combinatórias lexicais

Devido à complexidade envolvida, o elemento fraseológico das línguas é uma das maiores dificuldades para a tradução.

Conforme mencionado no subcapítulo 3.2, o grau de fixidez das CLs é um dos principais fatores que as distingue umas das outras. No entanto, independente do grau de fixidez, conhecer as CLs é fundamental para a tradução, pois: “O conhecimento das combinatórias mais ou menos fixas de palavras usadas numa língua é indispensável para que o tradutor seja capaz de produzir traduções que, para além de linguisticamente correctas, sejam naturais na língua de chegada” (Colaço, 2014). Assim, o estudo destas unidades é mesmo crucial para a tradução.

No subcapítulo 3.2, também falamos dos possíveis erros “*supressed slips*” que os falantes podem cometer ao utilizarem uma combinação lexical completa como “*cut my nails*” ao invés da que se pretendia utilizar “*brush my teeth*”, por exemplo. Na tradução, esta noção de unidade das CLs é fundamental. O tradutor deve ser capaz de identificá-las na LP e também deve utilizar expressões como um todo para traduzi-las, e não traduzir literalmente. Como afirma Jorge (1997a, p. 7), no caso das EIs: “Do ponto de vista tradutológico, a uma expressão corresponde uma outra expressão na outra língua, e pressupõe o seu conhecimento como um todo, como uma “palavra””.

Para a autora, reconhecer as CLs na LP é um desafio para o tradutor, sobretudo quando a LP não é a sua língua materna: “O falante apercebe-se da existência da lexicalização e da originalidade semântica de dada estrutura na sua língua materna, mas terá mais dificuldades em reconhecê-la e interpretá-la na língua estrangeira” (Jorge, 1997a, p. 7). Segundo a mesma, as dificuldades neste sentido incluem: 1) dificuldades de reconhecimento, pois uma EI pode confundir-se com frases não idiomáticas, 2) dificuldades de interpretação, quando o sentido

literal pode preceder o sentido idiomático e substituí-lo e 3) dificuldades de produção, quando o falante sente dificuldade em reutilizar a expressão num contexto.

Em relação às dificuldades de reconhecimento e interpretação, Costa (2017) afirma que são também pertinentes para o processo de anotação das colocações (classificação dos erros conforme uma categoria predefinida) e dá alguns exemplos. A autora menciona a dificuldade de reconhecer as CLs em diferentes variedades linguísticas<sup>4</sup>:

[...] surgem também problemas em reconhecer variação, nomeadamente usos no espanhol da América Latina que foram inicialmente marcados como incorretos, mas que num segundo momento, através da procura em *corpora*, foram também detetados usos em outros países de língua espanhola. Estes casos foram anotados como corretos, especificando a variante a que pertenciam (*tomar clases – ir a clase*). (Costa, 2017, p. 78)

Além de, por exemplo: “[...] colocações passarem despercebidas até surgir a sua versão incorreta. Por exemplo, *país de origem* não foi anotado como colocação até surgir a forma *países maternos*” (Costa, 2017, p. 95). O que reforça ainda mais a complexidade da tradução destas expressões.

Outro problema de tradução tem a ver com o sentido figurado que as CLs podem ter. Costa (2017) menciona, por exemplo, que, apesar de a transparência semântica possibilitar uma fácil decodificação das colocações a um aluno estrangeiro, por causa da possibilidade de terem um sentido metafórico, a produção das colocações não é igualmente fácil para este aluno. “Tratam-se de combinações onde existe um elemento que é usado num sentido metafórico ou “figurado”, convencionado, como *levantar* em *levantar dinheiro*” (Costa, 2017, p. 29). E isto

---

<sup>4</sup> Os problemas de tradução das CLs para diferentes variedades linguísticas são tratados na seção 3.3.3.

se aplica também à TA, pois o sentido figurado que algumas colocações podem empregar é sempre um desafio para os sistemas.

No caso da EI, principalmente, em que a metaforicidade é muito ampla, Vilela (2002, p. 176) afirma: “[...] não pode ser traduzida literalmente para outra língua, isto é, não é possível a tradução palavra por palavra, sem que essa expressão não tenha qualquer restrição, nem no plano sintático nem no plano semântico”. O autor afirma que, no caso das EIs, o sentido é opaco, não transparente e não composicional.

Assim, a tradução das EIs envolve dificuldades especiais:

O termo “idiomático” remete para o que é próprio de um idioma. De facto, as expressões idiomáticas refletem, em muitos casos, especificidades históricas e sócio-culturais de uma comunidade e daí advêm as dificuldades de encontrar os equivalentes diretos entre línguas. Estas idiossincrasias das línguas representam, sem dúvida, uma dificuldade acrescida para os tradutores, para os aprendentes estrangeiros, mas também para nativos, tratando-se de expressões marcadas por regionalismos. (Costa, 2017, pp. 16–17)

Mesmo no caso das colocações, em que pode existir uma transparência semântica, a tradução literal nem sempre é a mais ideal:

Sejam transparentes ou não, surge assim a definição de fraseologia por força da sua semântica própria: a combinação fixa de palavras que transporta um significado não composicional, ocasionado por recursos como a metáfora, a metonímia, a hipérbole, etc. (Vilela, 2002, p. 183)



Outro possível problema de tradução das colocações é que elas são diferentes nas diferentes línguas, podendo inclusive, ser colocação numa língua e na outra não. De acordo com McKeown & Radev (2000, p. 12), não se pode assumir que um conceito que é expresso através de uma colocação em uma língua utilizará também uma colocação em outra língua: “*Let us consider the English collocation to brush up a lesson, which is translated into French as repasser une leçon or the English collocation to bring about whose Russian translation is the single word verb осуществлять*” (op. cit.). Os autores também dão o exemplo *régler la circulation*, em francês: “*used to refer to a policeman who directs traffic, the English collocation. In Russian, German, and Serbo-Croatian, the direct translation of regulate is used; only in English is direct used in place of regulate*” (McKeown & Radev, 2000, p. 4).

Levar em conta todos estes problemas é especialmente importante para evitar a “destruição das locuções e frases idiomáticas”, das Tendências Deformantes de Berman (1985). É o que o autor propõe analisar com a chamada “analítica da tradução”: “o sistema de deformação dos textos – da letra – que ocorre em todas as traduções, impedindo-as de atingir o seu verdadeiro desígnio” (Jorge, 1997b, p. 41).

As tendências deformantes, além de estarem em todas as traduções, estão também em todas as línguas. Mas, apenas saber que elas existem não é suficiente para evitá-las, só uma “análise” da sua atividade permite neutralizá-las. É apenas ao submeterem-se a “controles” (no sentido psicanalítico) que os tradutores podem esperar libertar-se parcialmente deste sistema de deformação [...] (cf. Berman, 1985).

As perdas em tradução são inevitáveis, no entanto, saber identificar e neutralizar as tendências deformantes é importante para melhorar a qualidade das traduções, uma vez que isto permite ao tradutor respeitar as idiossincrasias do autor, mantendo na tradução as marcas do texto, através da mudança de estratégia sempre que necessário. Além disso, o conhecimento

das tendências deformantes nos permite desenvolver um senso crítico, evitando-nos cair nas “armadilhas” às quais estamos sujeitos no ramo da tradução.

A “destruição das locuções e frases idiomáticas”, tendência deformante que afeta diretamente as CLs, acontece quando não há a preocupação de manter os fraseologismos do texto original ou a sua carga fraseológica no texto de chegada (TC). Para evitá-la, a tradução não deve ser feita literalmente e não se pode por exemplo, em alguns casos, substituir referentes culturais que seriam incompreensíveis na LC por “equivalentes”, pois um referente cultural equivalente não mantém as idiossincrasias do autor. Berman dá como exemplo a EI de *Typhon* de Conrad: “[...] *Damne, if this ship isn't worse than Bedlam!*”, em que, “Bedlam” é um manicômio inglês. O autor argumenta que, neste caso, não seria adequada uma tradução literal da expressão, mantendo, por exemplo, a palavra “Bedlam”, que o público-alvo não entenderia. Por outro lado, na tradução para francês, por exemplo, também não seria ideal substituir “Bedlam” pelo equivalente em francês “Charenton”. Berman defende que mesmo que o sentido seja idêntico, substituir uma EI pelo seu “equivalente” é etnocentrismo que, repetido em grande escala, levaria ao absurdo de, em *Typhon*, as personagens se exprimirem com um monte de imagens francesas. Na tradução da EI deve-se, ao invés disso, respeitar a forma que a própria LC transmitiria aquela ideia naturalmente.

Por tudo isso, a tradução enquanto reflexão é altamente relevante. A literalidade não é possível na EI. A tradução de uma CL deve ser feita através de uma CL equivalente (se houver) ou da maneira que a LC expressaria aquele conceito naturalmente. Assim, a equivalência, por sua vez, envolve uma problemática que deve ser ponderada, cuidadosamente, na tradução das CLs. Pois a tradução destas expressões não consiste em ir à procura das palavras, mas da enunciação, das formas de dizer. Disso sim, deve-se fazer equivalência e é neste sentido que esta estratégia se torna possível na tradução das CLs (conforme veremos no subcapítulo 3.4).

De acordo com Jorge (1997b, p. 63), os conceitos de fidelidade e exatidão da tradução: “reportam-se à literalidade carnal do texto”. Isto é, as CLs carregam uma realidade carnal, tangível e viva ao nível da língua. “Enquanto perspectiva ética, o fim da tradução é acolher na língua materna esta literalidade” (*op.cit.*). Ou seja, na LC, as escolhas do tradutor também devem carregar essa realidade carnal, tangível e viva ao nível da língua, e, além de serem naturais para os falantes da língua, devem corresponder à variedade linguística com que o público-alvo se identifica culturalmente. Neste sentido, para evitar a destruição das locuções e das frases idiomáticas, é preciso também ter o cuidado de, por exemplo, não usar expressões do PE na tradução das CLs para o PB e vice-versa.

Nesta seção vimos que as tão complexas CLs são um dos maiores desafios para a tradução. Assim como na LP, a tradução deve também ser de expressões como um todo e não tradução literal. Alguns dos problemas de tradução que as CLs envolvem têm a ver com dificuldades de reconhecimento, interpretação e produção das mesmas, sobretudo quando não é a língua materna do tradutor ou mesmo variedade linguística; o sentido figurado, o fato de as línguas ou diferentes variedades de uma língua utilizarem diferentes CLs para expressarem o mesmo conceito, ou nem mesmo utilizarem uma CL naquele contexto, a problemática da equivalência, entre outros problemas, conforme veremos ao longo de todo este subcapítulo. E a análise destes problemas é crucial para evitar a referida destruição das locuções e frases idiomáticas no TC.

A próxima seção trata dos problemas de tradução das CLs especificamente no âmbito da TA.

### 3.3.1 Combinatórias lexicais e Tradução Automática

A TA, diferentemente da tradução assistida por computador, é feita unicamente pelos sistemas computacionais. Nesta última, o tradutor humano faz a tradução com o auxílio das ferramentas. No entanto, os resultados obtidos pela TA normalmente não são suficientemente bons para obter traduções com melhores níveis de qualidade, pelo que a pós-edição humana se torna crucial (cf. Testa, 2018).

Para falar dos problemas de tradução das CLs relativamente à TA, começaremos por fazer uma breve apresentação dos principais sistemas deste tipo de tradução até o momento, para então, na sequência, colocar os desafios envolvidos.

#### 3.3.1.1 Sistemas de Tradução Automática

A TA teve origem nos anos 50. A primeira década deu início à três abordagens básicas: a primeira era o modelo “tradução direta”, em que a tradução era de uma LP para uma LC específica, com o mínimo de análise linguística. A segunda abordagem era o modelo “interlíngua” baseado em representações abstratas de linguagem neutra (códigos ou símbolos, tanto para a LP quanto para a LC), em que a tradução seria então em dois estágios, da LP para a interlíngua e da interlíngua para a LC. A terceira abordagem era o modelo “*transfer*”, no qual a conversão era efetuada através de um estágio de transferência de representações abstratas (ou seja, não ambíguas) de textos da LP para representações equivalentes na LC. Neste caso, a tradução compreendia três etapas: análise, transferência e geração (ou síntese) (cf. Hutchins, 2015).

Baseados nos modelos “*transfer*” e “interlíngua”, nas décadas de 80 e 90 surgiram os sistemas de TA baseados em regras (***Rule-based Machine Translation (RBMT)***). Constistem na representação do conhecimento linguístico através de regras implementadas nos sistemas.

Nos anos 90, apareceram as “memórias de tradução”<sup>5</sup>. Com o avanço tecnológico, surgiram então, nas décadas de 2000 – 2010, os sistemas de TA baseados em *corpora*. O primeiro é o **SMT**, baseado em estatística. “*Statistical MT has different subgroups, including word-based, phrase-based [PBMT], syntax-based and hierarchical phrase-based*” (United Language Group, 2018). Tais sistemas utilizam *corpora* bilíngues anotados (chamados *treebanks*) conforme análise linguística e são feitos alinhamentos para encontrar as equivalências. Basicamente, “*Statistical MT is a “rule-based” MT method, using the basis of corpora translations to create its own text segments*” (idem).

A partir deste tipo de sistemas, foi possível começar a resolver o problema da ambiguidade, com a possibilidade de verificação de quantas vezes uma palavra foi traduzida de determinada forma. Mas, de qualquer modo, o SMT é melhor combinando o conhecimento linguístico com os estatísticos.

O segundo tipo de TA baseado em *corpus* foi o “*example-based machine translation*” (**EBMT**) (ou *memory-based*). Estes sistemas foram resultado de processos de extração e seleção de frases equivalentes ou grupos de palavras de um banco de dados de textos bilíngues paralelos que foram alinhados, ou por métodos estatísticos (que podem ser semelhantes aos utilizados na SMT), ou por métodos baseados em regras “*A major problem is the recombination of selected target language examples (generally short phrases) in order to produce fluent and grammatical output*” (Hutchings, 2015:14).

Posteriormente, combinando as abordagens estatísticas e as baseadas em regras surgiram os sistemas “**híbridos**”.

---

<sup>5</sup> “*A translation memory (TM) is a database that stores “segments”, which can be sentences, paragraphs or sentence-like units (headings, titles or elements in a list) that have previously been translated. The translation memory stores the source text and its corresponding translation in language pairs called “translation units”. Individual words are handled by glossaries (aka terminology bases) and are not within the domain of TM*” (Unbabel, 2018a, p. 5).

Por último, a fim de melhorar a precisão das traduções, surgem os sistemas **NMT**, um novo tipo de TA baseada em *corpora*. A NMT também utiliza métodos estatísticos, mas com utilização de aprendizagem profunda: “*NMT uses deep learning techniques to teach itself to translate text based on existing statistical models*” (United Language Group, 2018).

A denominação se dá pelo fato de que: “*NMT is based on the model of neural networks in the human brain, with information being sent to different “layers” to be processed before output*” (United Language Group, 2018). Apesar disso, segundo Forcada (2017), esta relação de NMT com os neurônios ou com a forma com que os cérebros das pessoas (ou dos tradutores) funcionam é muito vaga.

*The name comes from the fact that the neural networks (which should properly be called artificial neural networks) on which NMT is based are composed of thousands of artificial units that resemble neurons in that their output or activation (that is, the degree to which they are excited or inhibited) depends on the stimuli they receive from other neurons and the strength of the connections along which these stimuli are passed.* (Forcada, 2017, p. 292)

Esta nova abordagem de TA tem sido investigada intensamente nos últimos 4 ou 5 anos e tornou-se o atual estado da arte.

Como já referido, este é o sistema utilizado pela Unbabel e, no momento em que esta pesquisa foi feita, já era NMT genérica e adaptada aos domínios e clientes.

Em comparação com a SMT, a NMT traduz mais rapidamente e com melhor qualidade. “*NMT produces translations that are more accurate than its predecessor, Phrase Based Machine Translation (PBMT), thanks to its superior ability to translate complete sentences at a time*” (United Language Group, 2018).

Apesar disso, combinar a TA com o tradutor humano é crucial para garantir um alto nível de qualidade: “*The margin for error, no matter what method is used, is still too large to depend on MT for documents that will be published or used externally*” (*op.cit*).

### **3.3.1.2 Desafios das combinatórias lexicais para a tradução automática**

O estudo das CLs é fundamental para a TA, visto que estas expressões são muito desafiadoras para esta área da tradução, além de estarem presentes em todas as línguas e tipos de comunicação.

No subcapítulo 3.2, foi mencionada a característica “recorrência” das CLs. Esta propriedade é muito relevante para a TA, já que a linguística computacional utiliza técnicas estatísticas e *corpora* para construir um léxico bilingue, que pode ser utilizado como parte de um programa de TA (cf. McKeown & Radev, 2000). No entanto, de acordo com vários autores, como já referimos atrás, apenas a frequência de ocorrência das CLs não é suficiente para identificá-las.

Segundo Sanromán (2000, p. 171) “o facto de que dois lexemas co-ocorram frequentemente não é condição suficiente para existir uma colocação”. E como afirma Colson (2003, p. 59 *apud* Camacho, 2008, p. 51): “Idiomatismos, como um todo, costumam ser muito freqüentes em qualquer tipo de texto. Isso não significa, entretanto, que um idiomatismo específico aparecerá freqüentemente, mesmo em grandes *corpora*” (*op.cit*).

Fontenelle (1994, p. 47) acrescenta que, através do método estatístico, alguns pares de palavras que são extraídos por estes sistemas não são considerados colocações por alguns linguistas. O autor dá o exemplo dos pares *doutor* e *enfermeira*, que apresentaram muita frequência em alguns textos e indaga: São mesmo colocações? Ou apenas palavras que coocorrem no mesmo contexto por serem do mesmo campo semântico e seus significados estarem relacionados?

Além disso, no caso das colocações, há limitações na utilização das técnicas estatísticas por causa da distância colocacional, como vimos na seção 3.2.1. No capítulo 5, trataremos deste aspecto, propriamente, na análise do *corpus* desta pesquisa.

Corpas Pastor (2001, p. 100) aponta ainda para outras críticas que as técnicas estatísticas têm recebido, como: o fato de existirem colocações que são muito estáveis, mas pouco frequentes, pelo que podem não aparecer em um dado *corpus*; o fato de o enfoque estatístico não dispor de instrumentos para a análise semântica de uma determinada colocação; entre outras limitações.

Assim, de acordo com a autora, apesar de as ferramentas computacionais incorporadas nos programas de gerenciamento automático de *corpus* oferecerem informações preciosas ao pesquisador humano, este precisa filtrar os dados de acordo com diversos critérios, geralmente do tipo semântico. É preciso sempre analisar estas unidades em contexto: “*Aunque no se niega la existencia del significado léxico de las palabras, se propone analizar el significado en contexto, es decir, teniendo en cuenta las asociaciones de éstas en el sintagma, de donde proviene gran parte su significado global*” (Corpas Pastor, 2001, p. 98).

Como propõe Costa (2017), a utilização dos sistemas de extração de colocações é complementar a outras análises:



A utilização de cálculos de ocorrências, bem como uma perspectiva mais linguística podem ser utilizadas para a identificação de colocações, embora tenham diferentes implicações, nomeadamente na elaboração de um dicionário de colocações. Uma abordagem que tenha em consideração o número de ocorrências mas que também se baseie numa definição linguística pode ser uma abordagem. (Costa, 2017, pp. 35–36)

Por tudo isso, identificar as CLs através de valores estatísticos relacionados com a frequência de ocorrência é sempre um grande desafio. De maneira que a TA das CLs não é auto-suficiente e a revisão humana é indispensável.

Outro problema que as CLs implicam para a TA tem a ver com a interferência da LP. Relativamente aos erros de tradução de colocações, Costa (2017) faz uma comparação interessante entre um sistema de TA e um aluno aprendente, mostrando que a língua nativa (LP no caso do sistema de TA) interfere na LC, visto que não existe um domínio completo desta língua. Segundo Hashemi e Hwa (2014), até as semelhanças entre os dois tipos de produções são muito importantes, pois podem ajudar a prever padrões de erros e, desta forma, melhorar os sistemas de TA.

Nesta comparação, quanto à localização dos erros, Costa (2017, pp. 193–194) afirma que a maioria dos erros ocorreram no colocativo: “[...], por exemplo, traduzir *take a bath* por *\*ter um banho* em vez de *tomar um banho* ou traduzir *take a nap* por *\*tomar uma sesta* em vez de *fazer/dormir uma sesta*<sup>6</sup> [PE]” (idem). Em relação à descrição dos erros do estudo, os erros de “substituição” foram os mais comuns nos dois tipos de produção (dos alunos e das traduções automáticas). Consistem em: “[...] erros em que um elemento da colocação é substituído por uma palavra que não se adequa, mas que existe na L2 ou na língua alvo, no caso da tradução”

---

<sup>6</sup> Tirar um cochilo em PB.

(Costa, 2017, p. 195). A autora dá os exemplos: “\**dizer a história* em vez de *contar uma história* como tradução de *tell a story*, [...] ou \**terra nativa* em vez de *terra natal* para traduzir *native land*” (*op.cit*). De forma que a análise mostra claramente a influência da LP neste tipo de erro.

A interferência também pode ser da LC, no caso de diferentes variedades linguísticas, pois, conforme veremos adiante, o conhecimento de uma CL em determinada variedade linguística não garante a sua produção em outra variedade desta língua e, no caso de TA, sobretudo, os detectores automáticos de variedades linguísticas ainda apresentam muitas limitações, especialmente no âmbito das CLs, que envolvem muitos desafios em vários sentidos.

Por fim, os desafios de tradução de CLs no âmbito da TA também se aplicam ao processo de anotação<sup>7</sup> de erros. De acordo com Monti e Todirascu (2015), é comum os sistemas de TA traduzirem as unidades multi-palavras incorretamente (seja de abordagem estatística, baseada em regras, etc.). Ou seja, no processo de anotação, existem muitos erros envolvendo as CLs e o anotador deve ser capaz de anotá-los devidamente, visto que o processo de anotação impacta na qualidade e contribui para a melhoria dos sistemas de TA. Neste sentido, Costa *et al.* (2015) propõem uma taxonomia de erros linguisticamente motivada que, diferente das taxonomias tradicionais, engloba problemas das línguas românicas e não do inglês. Nesta taxonomia, os autores incluem, especificamente, os tipos “*collocational\_errors*” e “*Idioms*”.

Esta seção fez uma breve descrição dos sistemas de TA, cujo último avanço criou a NMT, sistema que combina métodos estatísticos com aprendizagem profunda, e apresentou os problemas que as CLs implicam para a TA, como as limitações da propriedade “recorrência”,

---

<sup>7</sup> O capítulo 4 descreve o processo de anotação da Unbabel e o capítulo 5 analisa-o, relativamente às CLs.

dos sistemas de extração de combinatórias e os possíveis erros por interferência da LP ou mesmo da LC, quando existem outras variedades linguísticas.

Até o momento, as CLs referidas são da língua comum. A próxima seção aborda os problemas de tradução das CLEs.

### **3.3.2 Combinatórias lexicais especializadas**

A fraseologia não está só na literatura, está na língua. Cada ciência tem a necessidade de criar palavras e precisa de terminologias fraseológicas.

Alguns autores defendem que as colocações, por exemplo, são fundamentais para identificar um domínio de linguagem especializada:

*Los estudios sobre los lenguajes de especialidad han puesto de relieve la importancia de las colocaciones como elementos caracterizadores de un registro determinado. Por ejemplo, las colocaciones de la palabra expediente con verbos como incoar, instruir, tramitar o archivar "declaran" abiertamente su pertenencia al registro jurídico y, dentro de éste, al ámbito judicial. (Corpas Pastor, 2001, p. 94)*

O problema é que o fato de existir uma fraseologia na língua especializada não a impede de ser polissêmica, ou seja, ter mais de um significado. Por exemplo, “teoria dos nós” existe na matemática, na marinha, na psicologia, na poesia, etc.

Além disso, uma palavra pode ser combinada com certas palavras em um domínio e com outras palavras em outro domínio. Por exemplo, o verbo “arquivar”, pode ser combinado com certas palavras na área jurídica, tal como “arquivar um processo”, bem como pode ser

utilizado em expressões da área da informática, por exemplo, “arquivar os jogos baixados da internet<sup>8</sup>”.

Mas as CLEs podem contribuir também para a desambiguação, justamente porque uma palavra costuma ser colocada com determinadas palavras em um sentido e com palavras diferentes em outro, possibilitando assim fazer a distinção das combinatórias nestes domínios: “*Thus bank might co-occur with river in one sense and savings and loan when used in its financial sense*” (McKeown & Radev, 2000, p. 2).

Assim, uma das vantagens de identificar uma CLE e seu domínio para a TA é justamente que essa identificação pode contribuir para desambiguar as palavras com mais de um sentido (muito presentes na língua comum, a propósito), facilitando a sua interpretação. No entanto, além de as CLEs poderem ser polissêmicas, estas expressões também envolvem a restrição combinatória. De forma que, na tradução, além da identificação dos domínios, é preciso levar em conta esta restrição.

Valente (2005) aponta para dois aspectos: “I. Há um número finito de UL [unidades lexicais] que gravitam em torno de uma palavra-chave” e: “II. A restrição da combinatória das UL do vocabulário de um discurso especializado reduz as possibilidades de escolhas nos processos de análise e síntese de um sistema de tradução automática” (Valente, 2005:184).

Como exemplo, a investigadora apresenta uma amostra do sub-domínio Navegação do Termináutica, que contém as possibilidades de CLs para “vento”, “atracar” e “embicar”. Neste exemplo, ela mostra que as opções de combinação lexical da palavra “vento” são restritas a um grupo de palavras (ex: pressão, força, ação, etc.), assim como para “atracar”, por exemplo, as opções se prendem com outro grupo de palavras (ex: navio, embarcação, jet ski, etc) e para “embicar”, outro grupo (ex: jet ski, lancha e ferry). E afirma que não há variação entre estes grupos, de maneira que: “[...] um «jet ski embica», mas que não é possível que um \*«navio

---

<sup>8</sup> Mais detalhes sobre este exemplo serão apresentados na seção 5.1.

embique»” (R. Valente, 2005, p. 184). Ou seja, o fato de as CLs pertencerem a um discurso específico não significa que isto facilite o processo de escolhas de tradução. As CLs são arbitrárias e ocorrem conforme o uso, e isto vale também para as CLEs.

Além disso, no subcapítulo 3.1 mencionamos a tendência para a utilização de verbos suportes ou verbos leves nas CLEs. As construções com este tipo de verbos não devem, por princípio, ser traduzidas literalmente e, por esta razão, frequentemente levantam problemas para a TA, por exemplo. Tal é o caso da expressão “pegar gripe” em PB, que tem como equivalente em PE “apanhar gripe”.

De acordo com L’Homme (2000, p. 91), as CLEs são definidas pela convencionalidade do uso, mas este uso é determinado por uma comunidade linguística específica, formada por especialistas da área. De forma que, para a tradução, além do conhecimento da língua e da identificação da área de especialidade, é necessário ter conhecimento dos usos típicos desta área. Sublinhado isto, são critérios de utilização das CLEs: “*aparición en el discurso (ocurrencia) y la de pertinencia para la aplicación (perfil de usuario)*” (Lorente, 2001, p. 16).

Em suma, as dificuldades de tradução das CLEs também estão ligadas ao reconhecimento e interpretação das CLEs na LP, devido à possibilidade de haver polissemia e, principalmente, à respectiva produção na LC, tendo em conta a restrição combinatória envolvida.

Esta seção apresentou algumas particularidades das CLEs, que devem ser consideradas para a tradução das mesmas. Pois, para fazer a tradução destas unidades numa área de especialidade, além do conhecimento das línguas de trabalho, deve-se ter acesso à terminologia destas línguas e, apesar de as CLEs tornarem mais fácil o seu reconhecimento, para produzi-las também deve-se levar em conta as restrições combinatórias envolvidas.

E as restrições cabem também às diferenças entre variedades linguísticas, pois as CLEs devem estar dentro da realidade linguística da cultura de chegada. A seção a seguir trata dos problemas relacionados com as variedades linguísticas.

### **3.3.3 Combinatórias lexicais e variedades linguísticas**

Traduzir não é encontrar equivalentes exatos. A língua é flexível, repleta de dialetos, regionalismos, etc. Os falantes têm as suas próprias idiossincrasias, que são diferentes nas diferentes culturas. Como diz Ortiz Alvarez (2009),

Sabe-se que todas as línguas têm os seus modos de dizer fixos, consagrados pela tradição e mantidos pelo uso. Aceitando-se que as comunidades tenham uma espécie de inventário de fórmulas conhecidas e usadas por seus membros como conhecimento coletivo, pode-se supor que diferentes comunidades tenham diferentes inventários. (Ortiz Alvarez, 2009, p. 27)

Além disso, o estudo das diferenças entre as CLs das variedades linguísticas ajuda a eliminar estereótipos, reduzir choques culturais, e proporciona uma reflexão sobre a própria cultura e uma compreensão da cultura do outro. De acordo com a autora,

[...] um discurso sobre a diferença, sobre as relações entre o Eu e o Outro, é sempre visto como algo ameaçador, que envolve atitudes de defesa e superioridade. No entanto, os elementos desta relação não devem ser vistos como pontos extremos como algo antagônico, oposto, mas como parâmetros que melhor podem elucidar a compreensão dos fenômenos sociais, da rede de inter-subjetividades. Como diz a

autora [Frias (1991)]: “a diferença não serve para atribuir uma identidade, mas para identificar”. (Ortiz Alvarez, 2009, p. 29)

Conforme mencionado anteriormente, a dificuldade em reconhecer as CLs passa pelas variedades linguísticas. Um dos maiores problemas aqui é a arbitrariedade. Nos exemplos de McKeown & Radev (2000), o inglês americano e o inglês britânico apresentam diferenças arbitrárias em expressões semelhantes: “*Thus, in American English one says set the table and make a decision, while in British English, the corresponding phrases are lay the table and take a decision*” (McKeown & Radev, 2000, p. 4).

Corpas Pastor (2001) dá também o exemplo de variações entre diferentes dialetos ou diferenças entre variedades linguísticas do espanhol: “*Por ejemplo, en español peninsular se prefiere el verbo dar como colocativo de opinión, mientras que la variedad chilena utiliza entregar una opinión.*” (Koike, 2000 *apud* Corpas Pastor, 2001, p. 94). De forma que as variedades possuem colocações próprias e típicas de suas culturas.

Também já foi mencionado que um dos problemas de tradução das CLs tem a ver com o sentido metafórico que elas podem ter. Isto afeta também a produção das CLs nas variedades linguísticas. O exemplo referido de Costa (2017) (“levantar dinheiro”, em PE) é uma das combinatórias mais recorrentes no *corpus* analisado no presente trabalho, no entanto, no PB, a colocação equivalente é “sacar dinheiro”, de modo que a combinação “levantar dinheiro” causa uma grande estranheza na variedade brasileira. Assim, o conhecimento das CLs do PE não possibilita a produção diretamente destas combinatórias no PB, e vice-versa.

A propósito, no caso das variedades linguísticas “PB e PE”, que são as variedades exploradas neste trabalho, muito já se tem falado na literatura sobre as diferenças gramaticais, contudo, não há muitos trabalhos sobre as diferenças no que tange as CLs. Mais uma razão

para o presente trabalho analisar as diferenças entre as duas variedades, essencialmente, aplicadas a estas expressões. A próxima seção trata, então, das duas variedades do português.

### 3.3.3.1 Português do Brasil e português europeu

O PB surgiu a partir do século XVI, quando o PE chegou ao Brasil e se misturou com as várias línguas indígenas brasileiras, várias línguas africanas (existentes devido ao tráfico negreiro) e, posteriormente, a partir do século XIX aproximadamente, a língua também entrou em contato com outras línguas europeias e não europeias que começaram a emigrar para o Brasil naquela época. “Este contacto foi configurando, através de uma história longa e complexa, o que é atualmente a variedade brasileira da língua portuguesa: o português do Brasil (PB), hoje falado por mais de cento e setenta milhões de pessoas” (Silva, 2013, p. 145).

Relativamente às diferenças gramaticais entre PB e PE, Silva (2013) sugere que as diferenças são “mais acentuadas quando se considera a variedade popular do PB e mais atenuadas quando se trata de falantes com alto grau de instrução que, em contextos altamente formais, procuram seguir a norma-padrão lusitanizante” (Silva, 2013:151).

No domínio da sintaxe, a autora menciona diferenças como a dos pronomes sujeitos (cujas diferenças afetam também o sistema da concordância verbal (em pessoa / número)) e os pronomes clíticos. Mória & Viotti (2004:115) apontam para a diferença em relação ao uso do gerúndio, usado mais frequentemente no PB: “*One of the most notorious differences between BP and modern EP [...] is that the latter resorts to infinitival forms, normally preceded by the preposition a (or por, in some cases), in many contexts in which the latter [former] uses gerundive forms*” (op.cit.). Outras diferenças sintáticas serão comentadas mais adiante, já no contexto das CLs.

Em relação ao léxico, segundo Silva (2013), é nesta área que estão as maiores diferenças entre as duas variedades.



[...] existe um fundo lexical comum e constante entre o português europeu e o português do Brasil, tendo o português europeu integrado algum léxico brasileiro indígena, como, por exemplo, jacarandá, piranha, abacaxi. Essa partilha lexical permite (com alguns mal-entendidos de permeio) uma comunicação não problemática entre brasileiros e portugueses, embora os usos discursivos, conversacionais e pragmáticos distingam bastante a interação linguística no Brasil e em Portugal. (Silva, 2013:154)

Estas diferenças no léxico das línguas também atingem, conseqüentemente, as CLs, conforme veremos na seção a seguir.

Ainda há muita controvérsia quanto ao estatuto do PB. Alguns autores defendem que a língua no Brasil deveria ser uma língua independente, “a língua brasileira”, enquanto outros autores a reconhecem como uma outra variedade da língua portuguesa (cf. Ortiz Alvarez, 2000). Independente disto, ainda há, de fato, muita confusão em relação à língua no Brasil, uma vez que há muitas diferenças, principalmente, no PB oral, nas formas de se comunicar, em comparação com o PE. Silva (2013, p. 153) menciona que o sistema pronominal do PB, por exemplo, gera problemas de insegurança linguística, uma vez que os falantes, “no processo de escolarização, são “corrigidos” em função do padrão da gramática prescritiva de tradição lusitanizante.”

Castilho (2010), já com esta preocupação, criou a chamada “Nova gramática do português brasileiro”, onde busca incorporar a oralidade na descrição gramatical. Esta nova abordagem tem como objeto de estudo a língua em seu estágio atual, seu contexto social e textual de uso. Trata do texto que deriva de uma conversação, ou seja, há toda uma preocupação com a interação, com o diálogo, que ocorre nos textos. Além disso, esta nova gramática, diferentemente das tradicionais, não parte de exemplos da língua literária, e sim da linguagem

corrente, partindo do princípio de que seus leitores são cidadãos comuns. E é uma gramática que trata unicamente do PB, colocando, assim, o problema da identidade.

Neste sentido, Bally (1961 *apud* Ortiz Alvarez, 2000:69) “acredita que para julgar o estado de evolução de uma língua não basta ler os autores contemporâneos. Torna-se necessário ouvir o povo, sentir o seu modo de falar, às vezes bastante diferente do que lemos nas páginas dos livros.” Portanto, analisar as diferenças entre PB e PE na perspectiva do uso real dessas variedades, (de que são um exemplo as CLs, que são definidas justamente pelo uso e não por aspectos gramaticais) é muito importante.

Assim, o foco da pesquisa desenvolvida no âmbito deste trabalho não são as diferenças meramente gramaticais entre PB e PE, estas já foram muito discutidas ao longo dos anos, mas sim as diferenças no contexto das CLs. Dito isto, a seção a seguir trata das diferenças entre PB e PE aplicadas às CLs.

### **3.3.3.2 Diferenças entre português do Brasil e português europeu aplicadas às combinatórias lexicais**

Além da escassez de trabalhos na área, o estudo comparativo entre PB e PE relativamente às CLs é também importante para a tradução pelo fato de que, no dicionário, há muitos significados de palavras que são possíveis em vários contextos diferentes, no entanto, o que o tradutor deve considerar é aquele que é adequado a cada contexto. “A tradução não é só uma equivalência de vocábulos ou expressões, mas é também uma equivalência de culturas, o que pressupõe conhecer todas as ressonâncias, presentes ou longínquas, das palavras e expressões” (Jorge, 1997a, p. 5). Por isso, o estudo das variedades linguísticas na perspectiva das CLs é crucial, as CLs expressam o desenvolvimento das línguas, cultura, identidade, etc.

Neste sentido, como sugere Castro (1996), o estudo comparativo entre PB e PE precisa ser feito sob várias perspectivas, para que não se afirme a diferença entre as variedades pela ausência destes estudos:

[...] para nos sentirmos todos mais à vontade neste domínio, que interessa tanto a brasileiros quanto a portugueses, seria conveniente desenvolver nos estudos linguísticos sobre o português uma prática de análises comparatistas em que as estruturas brasileiras e portuguesas fossem, por sistema, confrontadas, para não acontecer que se afirme a diferença na ausência, ou no desconhecimento, de um dos termos da comparação. (Castro, 1996: 139)

Assim, contribuímos aqui para essa nova abordagem dos estudos comparativos entre as duas variedades.

Já referimos que, de acordo com Silva (2013), as diferenças gramaticais são mais acentuadas quando se trata da variedade popular do PB e mais atenuadas quando se trata da norma culta. O mesmo pode ser afirmado nas diferenças em relação às CLs, pois quanto mais informal o contexto for, mais diferenças aparecem, mais idiossincrasias, etc. No entanto, como já mencionado, as CLs estão em todos os tipos de comunicação e podem aparecer, inclusive, em contextos formais.

Também já referimos a propriedade “arbitrariedade” das CLs e o seu impacto nas diferenças entre variedades linguísticas. Pois a escolha dos elementos utilizados em uma expressão é feita de acordo com a cultura, a comunidade de falantes e a forma como estes utilizam a língua. No caso do PB e PE, Camacho (2008, p. 134) mostra essa arbitrariedade com o seguinte exemplo: “a expressão brasileira “dormir sobre os louros”, onde o elemento para se caracterizar onde se descansa é o “louro”. Já na expressão portuguesa equivalente, o elemento

utilizado é outro: “dormir à sombra da bananeira””. Isto implica que as expressões seguem os *frames* (cenários), protótipos, estereótipos, etc. de cada cultura: “É com base na maneira como uma comunidade conceitualiza a realidade, que se escolhe o elemento a ser usado nas expressões” (idem).

Como concluiu Ortiz Alvarez (2000) no seu estudo da evolução do espanhol e do português na América, nem sempre as expressões de uma língua são semelhantes nas suas diferentes variedades: “Descobrimos que muitas das expressões cubanas são semelhantes às expressões brasileiras o que nem sempre acontecia com as expressões espanholas e as cubanas, ou com as expressões portuguesas e brasileiras” (Ortiz Alvarez, 2000, p. 68). Segundo a autora, a história das línguas influi também nas EIs, como é o caso das influências indígenas e africanas no PB:

O tupi legou ainda palavras que entraram nas expressões idiomáticas que com frequência usamos no dia-a-dia, por exemplo, andar na pindaíba, estar de tocaia, chorar pitangas. [...] O negro trouxe também sua contribuição à fraseologia brasileira representada por expressões como angü de caroço, chorar o lamba, banzé-de-cuia, virar ogó. (Ortiz Alvarez, 2000, p. 53)

Valente (2000) apresenta uma análise das diferenças e similaridades colocacionais entre PB e PE, de acordo com a TST de Mel’čuk, na qual, utiliza FLs que indicam os possíveis “colocativos” para as “palavras-chaves” (“bases” na designação de Hausmann (1985)), de acordo com o sentido que se pretende com a função. Por exemplo, para expressar “muito”, “intensamente” com a palavra “custar”, podem ser associados os colocativos “os olhos da cara”, “uma fortuna”. Esta análise está organizada em oito FLs da teoria, que conta com 56 funções no total: “Magn”, “Bon”, “Ver”, “Oper”, “Real”, “Incep”, “Fin” e “Caus”.

Em relação às três primeiras, a função “Magn” significa: “Intensificador [Magn]: modificador adjetival ou adverbial da palavra-chave P que expressa o sentido ‘muito’, ‘intenso’; a FL Magn liga um lexema a seu intensificador” (Valente, 2000, p. 10). A função “Bon”: “Laudativo [Bon]: modificador adjetival ou adverbial da palavra-chave P que exprime o sentido ‘bom’ enquanto aprovação subjetiva do locutor” (idem). E a função “Ver” é: “Confirmador [Ver]: modificador adjetival ou adverbial da palavra-chave P que expressa o sentido ‘tal como é preciso’” (idem).

Segundo a investigadora, as três funções são normalmente utilizadas com a FL “Anti”, denominada “Função Complexa” pela TST, que consiste na ligação de FLs simples, sintaticamente, cujo resultado é um sentido global, indecomponível. Por exemplo: “AntiMagn” (preço) = de banana, baixo.

As tabelas 4, 5 e 6 apresentam alguns dos resultados da análise da autora para estas três funções:

Tabela 4  
Magn (Extraído de Valente, 2000)

<b>entrada lexical</b>	<b>PE</b>	<b>PE &amp; PB</b>	<b>PB</b>
<i>custar</i> <sup>v</sup>	<b>AntiMagn2</b> <i>duas vezes nada</i>	<b>Magn2</b> <i>os olhos da cara; uma fortuna</i> <b>AntiMagn2</b> <i>nada; uma ninharia</i>	<b>Magn2</b> <i>uma facada; um absurdo; uma grana preta</i>
<i>preço</i> <sup>n</sup>	<b>AntiMagn</b> <i>da chuva; da uva mijona</i>	<b>Magn</b> <i>exorbitante; alto</i> <b>AntiMagn</b> <i>baixo</i>	<b>AntiMagn</b> <i>de banana</i>

Tabela 5  
Bon (Extraído de Valente, 2000)

<b>entrada lexical</b>	<b>PE</b>	<b>PE &amp; PB</b>	<b>PB</b>
<i>comida</i> <sup>n</sup>		<b>Bon</b> <i>//manjar (dos deuses)</i>	<b>Bon</b> <i>no ponto</i> <b>AntiBon</b> <i>//gororoba</i>
<i>dormir</i> <sup>v</sup>		<b>Bon</b> <i>como um anjo; como um bebê</i>	<b>Bon</b> <i>feito uma criança</i>

Tabela 6  
*AntiVer (Extraído de Valente, 2000)*

<b>entrada lexical</b>	<b>PE</b>	<b>PE &amp; PB</b>	<b>PB</b>
<i>falar</i> <sup>v</sup>	<b>AntiVer sem ser ouvido</b> <i>para o boneco</i>	<b>AntiVer sem ser ouvido</b> <i>para as paredes</i>	
<i>conversa</i> <sup>n</sup>	<b>AntiVer</b>	<b>AntiVer</b> <i>fiada; à toa; vazia; oca; sem pés nem cabeça</i>	<b>AntiVer</b> <i>pra boi dormir; de usineiro; mole;</i>

Em relação às demais funções: “a FL Oper formaliza a noção de verbo suporte (ou verbo leve). Essa FL e seus valores são verbos semanticamente vazios (ou esvaziados de sentido no contexto da palavra-chave)” (Valente, 2000, p. 11). Quanto à função “Real”: “a FL Real expressa, grosso modo, o sentido ‘realizar os “objetivos” inerentes ao referente designado pela palavra-chave’. Portanto, essa função se refere a verbos semanticamente plenos” (idem). As funções Incep & Fin, de acordo com Valente, não têm estrutura actancial própria e por esta razão elas se apoiam em outras que têm, como as funções “Oper” e “Real”, por exemplo. Segundo a mesma autora, “as FL Incep e Fin expressam duas fases diferentes de um estado ou acontecimento. Essas FL conhecidas como funções fásicas exprimem o sentido de verbos semanticamente plenos” (idem). Por fim, “a FL Caus exprime um dos tipos de “causação” de um estado ou acontecimento. Essa FL, conhecida como Função Causativa, exprime o sentido de verbos semanticamente plenos : Caus(P) = ‘causar que P [fazer de maneira que P aconteça]’” (Valente, 2000, p. 11).

As tabelas 7 e 8 apresentam mais alguns resultados da análise da autora para estas últimas funções:

Tabela 7  
*Oper (Extraído de Valente, 2000)*

<b>entrada lexical</b>	<b>PE</b>	<b>PB &amp; PE</b>	<b>PB</b>
<i>conversa</i> <sup>n</sup>	<b>IncepOper1</b> meter <b>IncepOper1</b> atracar		<b>IncepOper1</b> puxar
<i>marido</i> <sup>n</sup>	<b>Caus2Oper2</b> pescar [ART ~]	<b>Caus2Oper2</b> arranjar [ART ~]	<b>Caus2Oper2</b> fisgar [ART ~]
<i>moda</i> <sup>n</sup>		<b>FinFunc0</b> passar [de ~] <b>FinOper1</b> sair[de ~]	<b>FinFunc0</b> cair [de ~]

Tabela 8  
*Real (Extraído de Valente, 2000)*

<b>entrada lexical</b>	<b>PE</b>	<b>PB &amp; PE</b>	<b>PB</b>
<i>maconha</i> <sup>n</sup>		<b>Real1</b> fumar[~]	<b>Real1</b> puxar[~]

Todos estes exemplos comprovam que existem diferenças entre PB e PE até nas colocações. Estas FLs podem ser interessantes para a elaboração de dicionários, por exemplo. No presente trabalho, contudo, estas funções não serão utilizadas porque além de tratarem unicamente das colocações, conforme referimos Fontenelle (1994:6) na seção 3.2.1.2, as FLs focam nas relações semântico-lexicais básicas, como a sinonímia, antonímia e hiponímia. No caso da tradução das CLs, é preciso considerar que possuem restrições combinatórias e, portanto, as escolhas de tradução não devem ser feitas com base nos possíveis sinônimos, mas no uso típico da expressão.

Camacho (2008) faz uma análise sobre as diferenças entre as EIs do PB e do PE. A investigadora observa que, embora a maior parte dos idiomatismos brasileiros também seja usada em Portugal, há muitas diferenças entre os idiomatismos “quando estes descrevem as experiências cotidianas, os lugares, a identidade de uma nação” (Camacho, 2008, p. 135). Mostrando, assim, as várias formas de expressividade da língua.

A identidade de uma língua é construída nas diferenças, e as EIs ilustram essas diferenças, cujas múltiplas formas de expressão e variações a enriquecem. Em tempos de globalização e, conseqüentemente, de esvaziamento das fronteiras, essa afirmação da identidade torna-se necessária para resgatar o passado, a história e os valores culturais de um país. (Camacho, 2008, pp. 135–136)

Desta forma, a autora mostra como se constroem as diferenças e semelhanças entre as EIs do PB e do PE através dos seguintes parâmetros: 1) expressões semelhantes: traços quase universais, 2) expressões semelhantes formalmente, mas com sentidos diferentes, 3) expressões diferentes: traços não universais e 4) expressões diferentes: traços específicos.

Em 1), as expressões apresentam poucas diferenças na grafia, que, conseqüentemente, podem aparecer na fonética. O que ocorre, por exemplo, com a palavra “fato” (PB), que em PE é “facto”. Entram também nesta categoria diferenças na acentuação, como nas EIs “alma gêmea” (PB) e “alma gémea” (PE).

Ainda em relação às diferenças gramaticais, há as expressões que contêm o gerúndio. Como referimos Mória & Viotti (2004) atrás, os portugueses preferem o uso do verbo no infinitivo precedido pela preposição a, em detrimento ao gerúndio. Camacho (2008) dá como exemplo a expressão “de mãos abanando” (PB), que é equivalente à expressão “de mãos a abanar” ou “com as mãos a abanar” (PE). A autora também refere as expressões que contêm a preposição a indicando movimento: “os brasileiros geralmente usam a preposição em + a/o [...] Já os lusitanos utilizam a preposição a [...]”(Camacho, 2008, pp. 137–138). Segundo Camacho, isto se aplica também à preposição para, mais comum no Brasil “puxar brasa para sua sardinha”, cujo equivalente em PE é a expressão “puxar brasa à sua sardinha”.



Camacho (2008, p. 138) refere também a “clássica diferença da colocação dos pronomes oblíquos átonos”: no Brasil, principalmente na fala, tende-se a utilizar a próclise, enquanto que em Portugal, a ênclise é mais comum. Em sua análise, a autora mostra essas diferenças nas CLs: “Quando procurávamos as EIs brasileiras, colocávamos os pronome antes, como em “me atirei aos pés de”, “se colocou na pele”; já para a pesquisa das EIs portuguesas fizemos o contrário, colocando o pronome depois como em “atirei-me aos pés de”, colocou-se na pele”.

Além disso, foram classificadas na categoria (1), EIs idênticas, porém com variação de número. Por exemplo “bater as botas” (PB) no plural e “bater a bota” (PE); e a expressão “encher os olhos” (PB) e “encher o olho” (PE) que, no PE, são mais frequentes no singular.

Camacho (2008:139) encontrou ainda idiomatismos semelhantes, que contêm verbos diferentes, mas sinônimos. Exemplo “brigar com unhas e dentes” (PB) e “lutar com unhas e dentes” (PE). A autora afirma: “O mesmo processo de substituição acontece com a expressão brasileira “jogar na cara”, que significa “objetar algo a alguém”. Em Portugal, além dessa forma encontrada no Brasil, eles também utilizam a EI “atirar à cara””.

Por fim, outra curiosidade encontrada nesta categoria foi que, em relação às EIs brasileiras com o verbo “botar”, em PE o verbo é substituído pelo verbo “colocar” ou “meter”, uma vez que o verbo “botar” não é usado, pelo menos, na variedade padrão do PE.

No parâmetro (2), Camacho (2008) coloca as EI formalmente semelhantes, mas com um sentido diferente em cada variedade linguística. Isto acontece por causa do desenvolvimento natural das EIs, que sofrem mudanças e influências no decorrer do tempo. A autora dá o exemplo: “dar na cara”, que em PB significa “deixar transparecer algo” e “esbofetear alguém”, enquanto que em PE a EI tem apenas o sentido de “esbofetear”.

Em (3), que diz respeito à existência de EIs diferentes, Camacho (2008, p. 140) refere que este tipo de expressões “indicam a “ruptura” da língua portuguesa [...]”, já que “cada nação teve a sua história e sofreu influências diversas”. E explica:

Portugal, por ter sido uma grande potência marítima no século XVI e porta de saída da Europa para as Américas, teve contato com as mais diferentes culturas, o que, inegavelmente, trouxe mudanças à língua portuguesa. O Brasil, que já era habitado antes dos portugueses chegarem, recebeu, durante quase três séculos, uma imensa quantidade de imigrantes dos mais variados países. Portanto, seria impossível uma língua resistir a essas influências sem nenhuma mudança. (Camacho, 2008, pp. 140–141)

Como exemplos, Camacho menciona as EIs “entender do riscado”, que em PB significa “ser competente em certo domínio”, e que em PE corresponde a “saber da poda”.

Em relação às EIs que se referem à morte, existem três que são usadas nas duas variedades do português: “bater as botas” (a autora coloca esta expressão aqui, porém, como vimos atrás, no PE esta expressão é no singular), “esticar o pernil” e “ir desta para melhor”; porém, no PE, há ainda outras EIs que não são utilizadas no PB: “dar o berro”, “ir para o maneta” e “ir para os anjinhos”. Sobre a última, Camacho (2008, p. 141) afirma: “[...] parece relacionar-se com a expressão brasileira “ir para o céu”, geralmente usada para explicar a morte para uma criança”.

Para finalizar a categoria (3), a autora refere as EIs que “mantêm um léxico parcialmente parecido, mas que expressam imagens diferentes sobre o mesmo conceito” (Camacho, 2008, p. 141). Como exemplos, a autora menciona: “estar com a barriga roncando”

(PB), equivalente à “estar com a barriga a dar horas” (PE), que indicam “estar com muita fome”, segundo a mesma:

Os dois possuem a lexia “barriga”, mas isso não é o suficiente para que criem uma mesma imagem, pois as outras lexias que a acompanham ajudam a causar uma divergência de imagens. No caso da EI brasileira, o verbo “roncar” faz alusão a uma reação do organismo à fome, e, na EI portuguesa, o verbo “dar” e o substantivo “horas” constroem a imagem em cima do tempo que uma pessoa já está sem comer. (Camacho, 2008, p. 141)

No parâmetro (4), Camacho (2008) coloca expressões diferentes nas duas variedades linguísticas, com traços específicos de cada, como as encontradas: “expressões bastante curiosas que remetem à cultura portuguesa e que se utilizam de fatos históricos, geográficos, políticos e sociais de Portugal para tomarem forma” (Camacho, 2008, p. 142). Por exemplo: ““passar as passas do Algarve”, que seria o mesmo que “passar grandes dificuldades” ou, como se diz no Brasil, “comer o pão que o diabo amassou”” (*op.cit.*). Outros exemplos: “ser mais velho que a Sé de Braga” ou “ser do tempo dos Afonsinhos” em PE, expressões que valem-se de referências geográficas e históricas de Portugal para caracterizar algo muito antigo.

Segundo Camacho (2008), os portugueses dizem que “estão com os azeites” quando estão chateados. “[...] assim como, quando não estão para brincadeira vão logo dizendo para “ir chatear Camões” ou mandam “ir abaixo de Braga”, enquanto utilizamos no Brasil os equivalentes “ir catar coquinho” e “mandar às favas”” (Camacho, 2008, p. 142). (Esta última utiliza-se em PE no singular: “mandar à fava”). E no Brasil, existe a expressão “dar com os burros n’água”, quando alguma coisa dá errado, que em Portugal é “ficar em águas de bacalhau”.

Estes exemplos mostram a diversidade da fauna e da flora brasileira e os diferentes *frames* do Brasil e de Portugal. Pois o Brasil contém uma flora e fauna mais rica do que Portugal, logo, mais metáforas de imagem. Por exemplo: “ela é uma cobra / cascavel / serpente / jararaca”, entre outras expressões que mostram essa grande variedade de espécies. E, como já mencionado, as EIs são provenientes de expressões culturais.

Esta seção tratou dos problemas envolvidos nas diferenças entre PB e PE no âmbito das CLs para a tradução. Referimos que as escolhas de tradução são determinadas pelo uso e não apenas pelos significados existentes no dicionário, conforme a propriedade “arbitrariedade” das CLs também diz respeito às variedades linguísticas.

Vimos que, de acordo com Ortiz Alvarez (2000), nem sempre as expressões de uma língua são semelhantes nas suas diferentes variedades, já que a história, por exemplo, influi também nas EIs. De forma que uma expressão pode, por exemplo, ser parecida com uma língua diferente com história semelhante e diferente em outra variedade da mesma língua.

Referimos a análise de Valente (2000), baseada na TST de Mel'čuk sobre as diferenças e similaridades colocacionais e, por fim, mencionamos a análise de Camacho (2008) sobre as diferenças entre as EIs do PB e do PE.

A propósito, conforme mencionamos que as EIs são provenientes de expressões culturais, também o uso das CLEs é determinado pela cultura. A seção a seguir trata das diferenças PB e PE numa área de especialidade que apresentou muitas diferenças no *corpus* da presente pesquisa, a área da Informática.

### 3.3.3.3 Diferenças entre português do Brasil e português europeu nas combinatórias lexicais da informática

Com os avanços das tecnologias e a necessidade de comunicação cada vez mais rápida do mundo atual, a área da informática é caracterizada pela frequente inclusão de novos termos e expressões. Isto refletiu-se em uma abertura ao estrangeirismo nesta área na língua portuguesa.

Entretanto, a tendência para o estrangeirismo é notada mais frequentemente no PB do que no PE. Por exemplo, o termo especializado “rato” de Tecnologia da Informação (TI) em PE, no Brasil corresponde ao empréstimo *mouse*.

Além de o PB apresentar mais estrangeirismos, há outras diferenças em relação ao PE quando os termos são traduzidos. Bizarro (2014, p. 91), em sua análise sobre as diferenças tradutivas entre o PE e o PB num contexto específico de tradução técnica, argumenta que a “tradução técnica do PB tende a ser mais literal, parecendo haver uma maior propensão desta variante à assimilação, sobretudo de estruturas sintáticas e de terminologia do Inglês”. E complementa: “podendo mesmo classificar-se como decalque do Inglês: há uma adaptação do termo do TP à língua de chegada, sendo que esta adaptação acaba por ser uma espécie de estrangeirismo traduzido à letra” (Bizarro, 2014, p. 88).

A pesquisadora dá os exemplos: *(to) acess* (em inglês), que em PB é “acessar” e em PE é “aceder”, *(to) save*, que em PB é “salvar” e em PE é “guardar” e *connected* (inglês), que em PB é “conectado” e em PE “ligado”, entre outros. Estas diferenças entre os termos afetam, consequentemente, as CL da informática<sup>9</sup>. Por exemplo: “*to save the file*” (inglês) em PB é “salvar o arquivo”, enquanto que em PE a combinação é completamente diferente: “guardar o ficheiro”.

---

<sup>9</sup> Por “CLs da informática”, entende-se CLs que contêm termos da informática ou que são usadas frequentemente em textos desta área.

Os processos distintos de criação de palavras (em alguns casos, recorrer ao estrangeirismo e, em outros casos, traduzir os termos e expressões da área da informática<sup>10</sup>) ocorrem também no PE, porém em menor proporção, o PE tende a fazer mais adaptações de novos termos e expressões. Independente disto, esta pesquisa foca nas diferenças entre PB e PE no uso atual das variedades, já que, na tradução das CLs, é o uso que determina a escolha dos termos. Por esta razão, utilizar uma expressão do PE na tradução para o PB, por exemplo, por ser, talvez, uma escolha de tradução mais domesticante, ou seja, mais próxima da cultura-alvo, não é a solução adequada se o público-alvo, que é falante do PB, não a compreender (e vice-versa) ou se, nesta variedade, não for hábito usá-la.

Este subcapítulo apresentou os problemas de tradução, seus impactos para a TA, os problemas que envolvem as CLEs e, por último, os problemas relacionados com as variedades linguísticas.

Para concluir o capítulo, a seção a seguir aborda um último tópico, as estratégias de tradução das CLs.

### **3.4 Estratégias de tradução das combinatórias lexicais**

Ao longo deste capítulo, com o levantamento dos problemas de tradução das CLs, foi mencionado que estas expressões não devem, por via de regra, ser traduzidas literalmente e foi mencionada a problemática da equivalência, que, no caso das CLs, deve ser equivalência da enunciação, das formas de dizer. Foi também referido que a equivalência deve respeitar a variedade linguística. Neste sentido, a equivalência é uma estratégia de tradução das CLs, sobretudo das CLEs, onde não há criatividade, que tem que ser tradução etnocêntrica.

---

<sup>10</sup> Às vezes, os dois processos podem ser utilizados para referir o mesmo termo. Por exemplo, “e-mail” e “correio eletrônico”.

Esta seção apresenta as estratégias de tradução propostas na literatura de acordo com Misri (1990 *apud* Jorge, 2002) e a crítica que o autor faz em relação às estratégias existentes, mostrando quais elementos devem, então, ser levados em consideração na tradução das CLs.

O autor afirma: “As propostas apresentadas na literatura sobre a tradução da fraseologia podem resumir-se aos seguintes aspectos”: (i) tradução por um equivalente preexistente; (ii) tradução elaborada a partir de uma equivalência de situação, (iii) tradução de palavra a palavra, com nota; e (iv) tradução por equivalente preexistente, com nota. (cf. Misri, 1990 *apud* Jorge, 2002:5).

De acordo com o autor, (i) pressupõe que já existe um equivalente na LC. A estratégia (ii) é usada quando não existe um equivalente direto à expressão na LC, mas existe uma situação equivalente: “Neste caso, perserva-se o sentido da expressão, mas neutraliza-se o valor idiomático do texto” (Misri, 1990 *apud* Jorge, 2002, p. 7). No caso de (iii), a solução é compartilhada pelo teórico Antoine Berman<sup>11</sup>, e tenta manter o caráter exótico do texto original, privilegiando a palavra a palavra. Neste caso, utiliza-se a nota para facilitar a compreensão, porém, isto pode tornar a leitura muito pesada e menos espontânea. Por fim, utilizando a estratégia (iv), “privilegia-se a inteligibilidade do texto de chegada, tal como na primeira solução, mas favorece também a riqueza do texto de partida, pois a nota daria conta da tradução literal da expressão do texto original” (Misri, 1990 *apud* Jorge, 2002:8).

Após expor as estratégias mais comuns de tradução das CLs, Misri critica estas estratégias, mostrando que os autores se baseiam sempre no nível da língua e não no do discurso: “As expressões, bem como qualquer tipo de fraseologia, deveriam situar-se ao nível do discurso. É aí que elas ganham a sua própria autonomia, quando inscritas num acto de

---

<sup>11</sup> “Berman afirma, preservar a lexicalização da língua de partida não corresponde a um mero exercício de literalidade, mas à perservação de traços inerentes ao próprio processo de lexicalização e à construção das fraseologias.” (Jorge, 2002:6)

comunicação” (idem). Ou seja, as estratégias apresentadas não dão conta dos valores pragmáticos.

Assim, o autor indaga qual é o papel do tradutor na tradução deste tipo de estruturas e sugere:

Em todo o acto de tradução, seja ele idiomático ou não, o papel do tradutor é de dar conta e de interpretar o que o autor exprimiu e de transpor para uma outra língua, exprimindo nela os mesmos efeitos que sentiu aquando da leitura do texto original.  
(Misri, 1990 *apud* Jorge, 2002, p. 8)

Desta forma, para o autor, a tradução deve levar em conta quatro “componentes”: (1) componente informativa, (2) componente hierárquica, (3) componente colocativa e (4) componente de conformidade.

Segundo o autor, a componente (1) “prende-se com a mensagem informativa que advém da carga comunicativa contida na expressão, isto é, o que é restituído pela sua paráfrase” (Misri, 1990 *apud* Jorge, 2002, p. 8). Exemplos: “falar a torto e a direito” - falar muito e de qualquer assunto, sem discernimento; e “dar pontapés na gramática” - cometer erros.

A componente (2) “Prende-se com os níveis de língua. Embora várias expressões possam ser parafraseadas da mesma maneira, elas não são sinónimos e distinguem-se pela pertença a níveis de língua diferentes” (*op.cit*:9). Jorge dá os exemplos: “ter o dom da palavra”; “não ter freio na língua”; “ter paleio” e “não ter papas na língua”, em que todos dão conta da mesma interpretação geral (mesma componente informativa) “falar muito”, no entanto, não constituem verdadeiros sinónimos. Ou seja, essa componente tem a ver com a competência do tradutor para seleccionar a expressão adequada, por exemplo, ao registo do texto.



A componente (3) “propõe, ou defende, a tradução de uma lexicalização por outra lexicalização na outra língua.” (Misri, 1990 *apud* Jorge, 2002, p. 9). A componente (1) (paráfrase) não deve ser aplicada em todos os casos, porque a expressão perde o caráter lexicalizado (e, por isso, idiomático) do texto, favorecendo a neutralização idiomática em detrimento da expressividade (cf. Jorge, 2002). Assim, sempre que possível, o tradutor deve optar pela utilização de uma expressão equivalente na LC.

Finalmente, na componente (4), “A conformidade prende-se com os usos próprios de cada língua. [...] recusa-se a proposta de Berman da palavra a palavra, pois introduz na língua de chegada estruturas que, embora gramaticalmente correctas, não correspondem aos hábitos linguísticos dos falantes” (Misri, 1990 *apud* Jorge, 2002, p. 9). O autor afirma, assim, que, por vezes, existem várias expressões possíveis na LC para a tradução de determinada expressão na LP, por exemplo, e a tarefa do tradutor é, então, escolher a que mais se adegue à LC, pois as línguas têm maneiras diferentes de exprimir estados, emoções, ações, etc.

Portanto, é preciso analisar quais são as necessidades de cada tradução e a tradução das CLs deve ser da enunciação, não de palavras. No caso do *corpus* analisado no presente trabalho, especialmente porque os textos de partida são e-mails, tíquetes de suporte, etc., ou seja, conteúdos dinâmicos, que têm como objetivo, por isso, a comunicação rápida e clara, e estão repletos de valores pragmáticos, a tradução deve ser mais orientada ao nível do discurso, levando em conta as componentes sugeridas por Misri. E é justamente por isso que as CLs na LC devem ser típicas da cultura de chegada.

Independente da estratégia, as perdas semânticas das CLs da LP são inevitáveis: “a tradução proposta normalmente não recobrirá na totalidade o sentido do termo da outra língua” (Darbelnet, 1970 *apud* Camacho, 2008, p. 36). Mas o importante é que a tradução destas expressões deve ser realizada de acordo com o contexto de uso, a tradução literal nem sempre

é a melhor solução. É isto o que deve ser considerado para melhorar a qualidade das traduções provenientes de TA, para que estas sejam fluentes na LC. E isso também se aplica ao processo de pós-edição deste tipo de tradução.

Neste capítulo, foi apresentado o estado da arte do fenômeno “Combinatórias Lexicais”, na primeira metade do capítulo, foram exploradas as propriedades, os tipos e toda complexidade que estas expressões envolvem. Na segunda metade do capítulo, foram tratados os respectivos problemas de tradução. Normalmente, dificuldades de reconhecimento, interpretação e produção, pois a tradução deve ser da expressão no seu todo e não tradução literal. Foram explorados os impactos, sobretudo, para a TA, para as diferenças entre variedades linguísticas e para as CLEs. Por último, foram apresentadas as respectivas estratégias de tradução, que, no caso das CLs, devem ser consideradas ao nível do discurso e levar em conta os aspectos pragmáticos.

No próximo capítulo, apresenta-se a metodologia de trabalho utilizada.

## 4 Metodologia

A metodologia deste trabalho segue as seguintes etapas: em primeiro lugar, a seção 4.1 apresenta o processo de anotação da empresa. A tipologia é apresentada na seção 4.1.1; a métrica de qualidade na seção 4.1.2; finalmente, os critérios de anotação são apresentados na seção 4.1.3. Em seguida, a seção 4.2 descreve o *corpus* utilizado e por último, a seção 4.3 trata do glossário que foi criado.

### 4.1 Processo de anotação

O processo de anotação da Unbabel está a cargo do departamento de qualidade da empresa. Como já foi referido atrás, este processo consiste na anotação dos erros subsequentes da pós-edição. É feito por anotadores humanos através da plataforma “*Annotate*” e serve para avaliar a qualidade de TA e de *human-quality (HQ)*, domínios, etc. A classificação dos erros é feita com base numa tipologia de erros e níveis de indicação da qualidade preestabelecidos, conforme veremos adiante.

Os anotadores, linguistas profissionais, devem seguir as *Guidelines* de anotação e têm acesso a várias ferramentas para informações adicionais sobre o texto traduzido, conforme já referido. Todas essas facilidades são muito úteis e importantes para os anotadores, entretanto, a constante implementação de ferramentas requer o *feedback* dos usuários para que estas ferramentas sejam sempre melhoradas.

Em relação às CLs, no período em que o estágio foi feito, este processo não tratava especificamente deste tipo de unidades. Não existia um tipo de erro específico, tampouco critérios de anotação para estas expressões, o que motivou a presente pesquisa. Na próxima seção, mostraremos, então, os tipos de erros implantados na taxonomia da empresa e em quais deles as CLs tendiam a ser anotadas.

### 4.1.1 Tipologia

A figura 6 mostra a tipologia implementada no sistema, no período em que este estudo foi realizado.

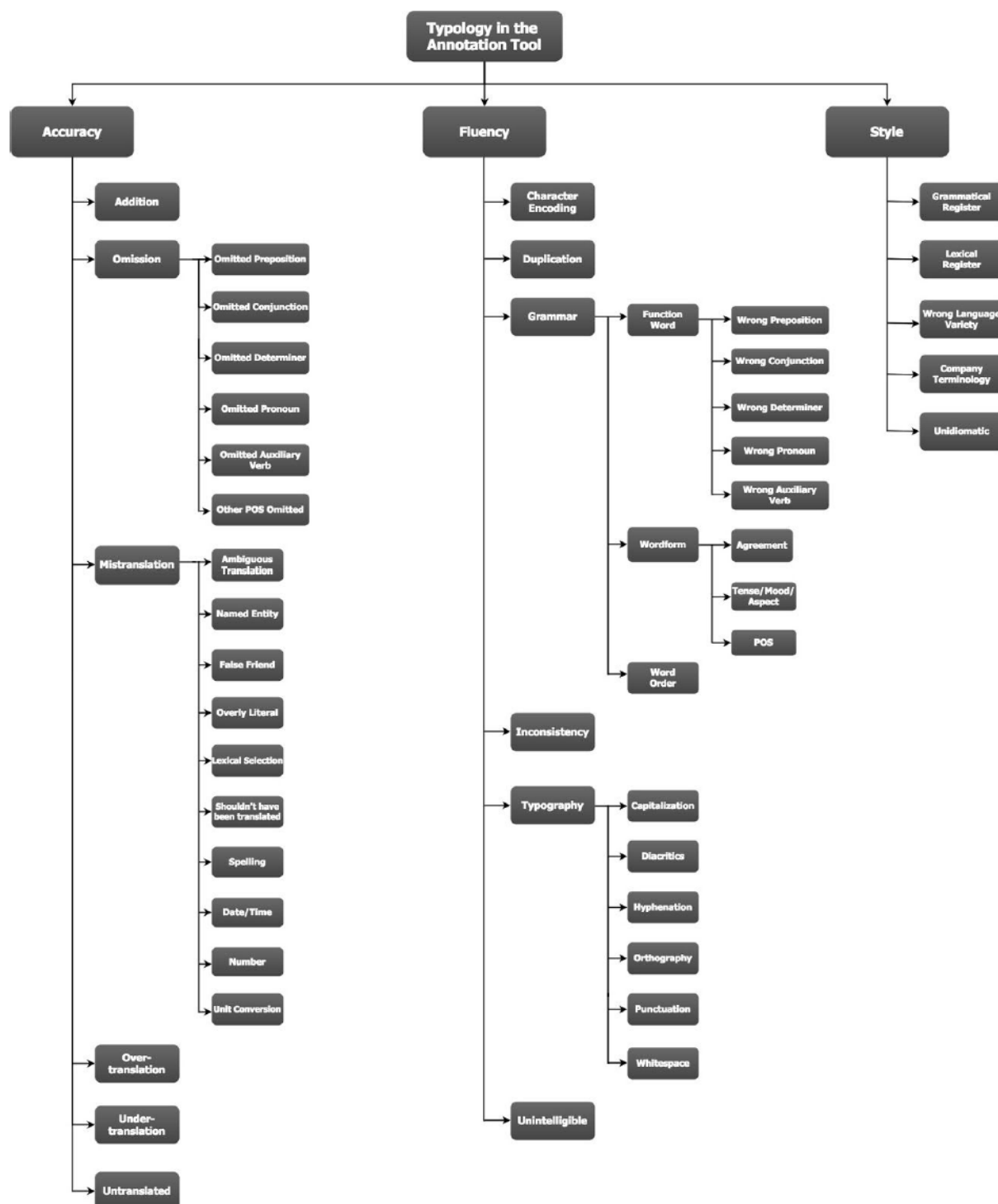


Figura 6. Tipologia de erros do sistema de anotação. (Extraído de Unbabel, 2018b, p.15)

A tipologia apresentada na figura 6 foi inspirada na métrica de qualidade que será descrita na seção seguinte e adaptada conforme as necessidades e o tipo de texto que a Unbabel traduz. A classificação de erros ocorre em relação a três quesitos cruciais da qualidade: “*Accuracy*”, “*Fluency*” e “*Style*”. O nível “*Accuracy*” (precisão) está relacionado com o sentido na LC, “*Fluency*” está ligado com a qualidade do TC, se este está bem escrito e inteligível e “*Style*” com o estilo do texto, registro utilizado, se está de acordo com glossário ou instruções do cliente, etc. A classificação correta de erros é fundamental para assegurar a qualidade das traduções.

Conforme mostra a figura 6, não há um tipo de erro específico para as CLs nesta taxonomia. Isto é um problema, pois, as CLs podem ser associadas a diversos tipos de erros, quando não existe uma categoria específica para elas. Com esta preocupação, este trabalho analisa as CLs em três tipos de erro que estão “mais próximos” dos problemas que ocorrem com estas expressões, além de estarem entre os tipos de erros mais frequentes: “*Overly Literal*”, “*Lexical Selection*” e “*Wrong Language Variety*”.

As instruções sobre a utilização de cada tipo de erro estão disponibilizadas nas “*Guidelines*” do processo de anotação da empresa. Neste trabalho, trataremos exclusivamente dos três tipos de erros que analisamos.

De acordo com as *Guidelines*, o tipo “*Overly Literal*” deve ser selecionado quando: “*The translation is too strict to the source text, which may cause problems of interpretation (like the literal translation of idiomatic expressions)*” (Unbabel, 2018b, p. 17). O tipo “*Lexical Selection*” deve ser utilizado quando: “*The term selected is not correct for the context or is not accurate to convey the meaning of the original text*” (idem). Por exemplo: “Complacente com a nossa política” em vez de “Compatível com a nossa política (escolha errada do adjetivo). Observe-se que esta descrição trata da anotação de termos, não menciona a possibilidade da anotação de, por exemplo, uma expressão. Apesar disso, para este tipo, as *Guidelines* orientam

para que se confirme se o termo selecionado não se enquadra no tipo de erro “*Overly Literal*” antes de proceder à anotação. Ou seja, os dois tipos podem ser confusos para o anotador. Finalmente, o tipo de erro “*Wrong Language Variety*” deve ser utilizado quando a variedade linguística utilizada está incorreta. Por exemplo, o uso do inglês britânico ao invés do inglês americano, do PE ao invés do PB e vice-versa, entre outras variedades linguísticas. Mais detalhes sobre as *Guidelines* são expostos na seção 4.1.3.

A próxima seção apresenta a métrica utilizada para o processo de anotação.

#### 4.1.2 Métrica

A métrica de qualidade utilizada na empresa é inspirada no modelo “*Multidimensional Quality Metrics*” (MQM). Trata-se de uma métrica desenvolvida originalmente através do projeto europeu “*QTLaunchPad*” que oferece uma lista flexível de tipos de erros. “*QTLaunchPad was a European Comission-funded collaborative research initiative (2012-2014) dedicated to preparing the grounds for a new type of systematic research and cooperation between MT developers and language industries for overcoming quality barriers[...]*” (Unbabel, 2018a).

Além de possibilitar à empresa determinar a qualidade do TC através da identificação de erros, esta métrica é também uma ferramenta útil para a implementação de melhorias, já que a empresa pode verificar, por exemplo, quais são os erros mais frequentes e investigar possíveis soluções.

Os anotadores devem selecionar o tipo de erro e classificar o erro conforme os níveis de indicação do impacto na significância e na fluência do TC em “*minor*”, “*major*” ou “*critical*”. A figura 7 mostra como os níveis devem ser indicados dentre os tipos de erro.

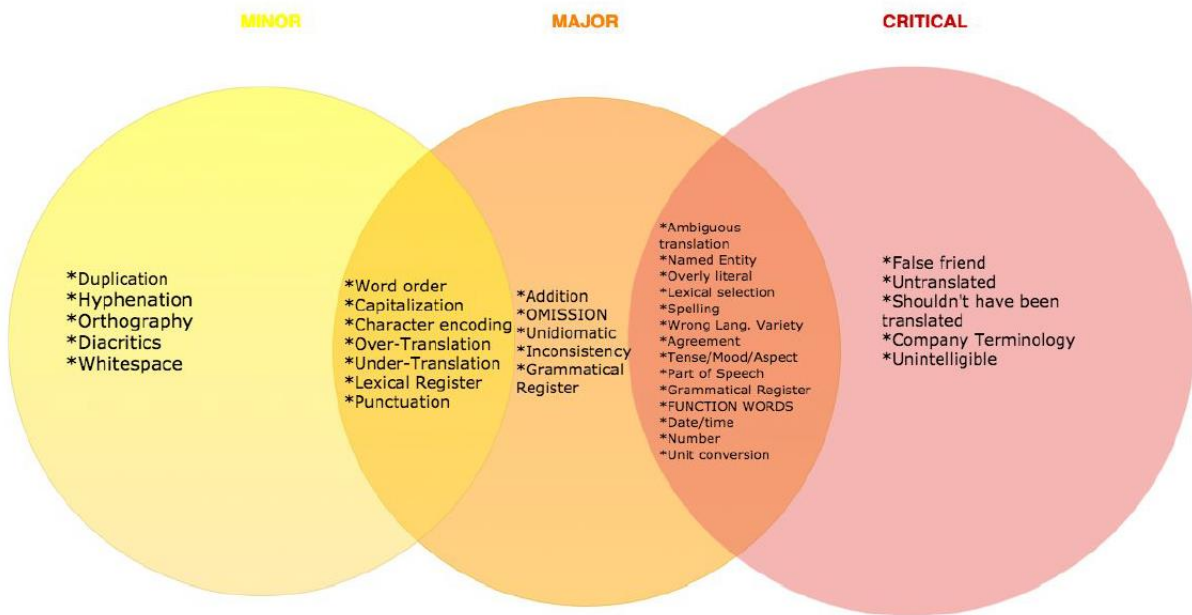


Figura 7: Three severity levels and intersections for type of errors. (Extraído de Unbabel, 2018b, p. 24)

Conforme mostra a figura 7, alguns tipos de erros devem ser classificados como *minor*, *major* ou *critical*, enquanto que outros ficam a critério do anotador decidir se são *minor* ou *major*, ou *major* ou *critical*, consoante o contexto, os exemplos a anotar e as indicações específicas dos clientes.

Após classificar o erro em um dos três níveis, o anotador deve, ainda, indicar a fluência em uma escala de 1 a 5, onde 1 indica o menor nível de fluência e 5 o maior nível. No modelo MQM original, são atribuídos os pesos: 1 para o nível “*minor*”, 10 para “*major*” e 100 para “*critical*”. Com isto, a métrica atribui a qualidade através da seguinte fórmula: “ $TQ = 100 - TP + SP$ ”, na qual  $TQ$  = *quality score* (a avaliação geral da qualidade),  $TP$  = *penalties for the target content* (soma dos valores atribuídos ao TC) e  $SP$  = *penalties for the source content* (soma dos valores atribuídos ao SC) (cf. Lommel, Burchardt, & Uszkoreit, 2015).

Apontada a métrica de qualidade, passemos para os critérios de anotação da empresa.

### 4.1.3 Critérios de anotação

Os critérios de anotação estão disponibilizados nas *Guidelines* do processo de anotação da empresa. Trata-se de um documento que apresenta o sistema de anotação e a tipologia de erros utilizados pela empresa, e dá instruções sobre como utilizar o sistema, como selecionar cada tipo de erro, com exemplos de cada erro, e como avaliá-lo de acordo com a métrica de qualidade adotada.

De acordo com as *Guidelines*, a unidade mínima que o anotador pode anotar é uma palavra e a máxima é a frase ou expressão toda. Contudo, espaços brancos, palavra faltando ou pontuação também podem ser anotados. Mas, apesar da possibilidade de anotar expressões inteiras, assim como no sistema de anotação não há um tipo de erros específico para as CLs, as *Guidelines* também não tratam especificamente deste tipo de unidades.

Isto é um problema, pois na tradução, o conhecimento das CLs (e dos problemas envolvidos) é indispensável para o reconhecimento das mesmas na LP e para a respectiva produção na LC. Assim, o anotador também deve ter conhecimento destas unidades e suas implicações para a TA (inclusive para a pós-edição), e ser capaz de identificar se a tradução está linguisticamente correta e natural na LC para fazer a anotação adequadamente.

Aliás, o processo de anotação dos erros serve também para a implementação de melhorias com base nos erros que foram anotados. Neste sentido, a falta de critérios relativamente à anotação das CLs pode resultar em anotações inconsistentes e insuficientes destas expressões (conforme mostramos no subcapítulo 5.2) e, consequentemente, refletir na qualidade.

Assim, a forma de anotação do erro é altamente relevante no caso das CLs. Por esta razão, as *Guidelines* poderiam dar instruções sobre como fazer a anotação das CLs, dado que uma melhor anotação dos erros reflete numa melhor resolução dos mesmos.



Além disso, as *Guidelines* para a anotação não existem para cada língua de trabalho, como ocorre com as *Guidelines* para a pós-edição. Em vez disso, estão em inglês e as regras valem para as demais línguas.

No âmbito das CLs, isto também pode ser problemático, devido à dificuldade de reconhecimento das CLs pelo anotador, acrescida quando não é a sua língua materna. A identificação deste tipo de erros é muito desafiadora para o anotador quando não há critérios para isso com os exemplos na sua língua nativa. Embora os problemas de tradução das CLs sejam os mesmos para as todas línguas, as suas propriedades podem se manifestar em maior ou menor grau dependendo da língua, de forma que os problemas também podem ocorrer em maior ou menor grau, dependendo da língua. Além de, como já foi referido atrás, por serem arbitrárias, as CLs apresentarem diferentes restrições combinatórias e diferentes formas de ganharem sentidos metafóricos nas diferentes línguas, de forma que pode ser difícil associar exemplos do inglês com outras línguas.

O conhecimento das CLs reflete um elevado nível de proficiência e, considerando que os anotadores podem trabalhar com outros pares de línguas que não envolvem o inglês, não ter este nível de inglês pode, então, ser um impasse. Mas, mesmo nos casos em que têm, evidentemente as CLs são sempre desafiadoras para a língua estrangeira, pelo que exemplos da língua nativa do anotador são muito úteis na tarefa de anotação dos erros.

Além disso, como referem McKeown & Radev (2000), nem sempre um conceito que é expresso através de uma CL em uma língua utilizará também uma CL em outra língua, o que também dificulta muito a associação de exemplos de expressões em inglês com equivalentes em outras línguas.

Neste sentido, as *Guidelines* para o processo de anotação, pelo menos em relação às CLs, poderiam ser específicas para cada língua de trabalho, dando exemplos de erros e soluções

de cada língua, evitando assim a falta de clareza, pois a utilização das CLs não é uma questão de utilização de recursos estilísticos, é uma questão de precisão.

Nesta seção, apresentamos o processo de anotação da empresa, o sistema “*Annotate*”, a tipologia de erros, a métrica utilizada e os critérios de anotação. A próxima seção faz a descrição do *corpus* analisado.

## 4.2 Descrição do *corpus*

O *corpus* recebido inicialmente foi baseado nos erros mais frequentes de tradução de inglês para PB de acordo com o processo de anotação da empresa e é composto por: data da anotação, breve descrição do cliente, tipo de erro, nível, erro (anotado), segmento na LP e segmento na LC.

Os tipos de erros compreendidos eram: “*Lexical Selection*”, “*Overly Literal*”, “*Punctuation*”, “*Untranslated*” e “*Capitalization*”. No entanto, os três últimos foram descartados, já que, como já referido, foi nos dois primeiros tipos que foram encontrados os melhores potenciais de investigação para o tópico deste trabalho.

Além de buscar reduzir estes erros mais recorrentes, optou-se por analisar também o tipo de erro “*Wrong Language Variety*”, pois estava diretamente relacionado com o tema da pesquisa. Desta forma, foi recebido um segundo *corpus* que, além de “*Wrong Language Variety*”, incluía os seguintes tipos de erros: “*Omitted Preposition*”, “*Wrong Preposition*”, “*Omitted Determiner*”, “*Omitted Pronoun*” e “*Wrong Pronoun*”.

Assim, o primeiro e o segundo *corpora* recebidos totalizavam oito tipos de erros (sem contar os tipos já descartados do primeiro *corpus*), cujos erros foram anotados em um período de nove meses, de 1 de junho de 2017 a 1 de março de 2018.

Para efeitos de proteção da privacidade dos clientes, todos os dados foram devidamente anonimizados, conforme o novo regulamento de proteção aos dados determina (o chamado Regulamento Geral da Proteção de Dados (RGPD)). A anonimização foi feita caso a caso, através de etiquetas, substituindo os dados dos clientes por etiquetas genéricas, também chamadas na literatura de “*placeholders*”. O procedimento realizado foi classificar os dados em grandes grupos como: “NAME” para nomes de pessoas, “COMPANY” para nomes de empresa, “PRODUCT” para nomes de produtos, etc. Além disso, considerando que a LC é o português, foram adicionadas informações de gênero e número, por exemplo “NAME-FEMALE”, “PRODUCT-SINGULAR”, quando necessário.

A tabela 9 mostra as quantidades de segmentos de erros recebidas por cada tipo.

Tabela 9  
*Quantidade de segmentos recebidos*

Type	
Lexical Selection	282
Overly Literal	127
Wrong Language Variety	65
Wrong Preposition	23
Omitted Pronoun	14
Omitted Preposition	9
Omitted Determiner	8
Wrong Pronoun	2

Considerando estas quantidades, optou-se ainda por utilizar neste trabalho apenas os três tipos de erros com mais segmentos de texto com erros, respectivamente “*Lexical Selection*”, “*Overly Literal*” e “*Wrong Language Variety*”, que, além de apresentarem mais ocorrências de erros, conforme mencionado anteriormente, envolvem muitos problemas relacionados com as CLs, que são o foco deste trabalho.

Estes três tipos de erros também são relevantes porque causam um grande impacto na qualidade das traduções. De acordo com as orientações das *Guidelines*, os três tipos são classificados normalmente como “*major*” ou “*critical*” no nível de indicação do impacto na significância e na fluência do TC.

O presente trabalho foi feito inteiramente com base no *corpus* recebido, aqui descrito, sem acesso a ferramentas de informações adicionais a que os revisores, por vezes, têm acesso, conforme referido no capítulo 2.

Após delimitar o *corpus* pelos três tipos selecionados, fizemos uma anotação do *corpus* no que diz respeito aos erros de CLs. Assim, dos três tipos analisados, que totalizam 474 segmentos de texto, foram encontrados 330 erros de tradução das CLs, em 286 segmentos. Os 188 segmentos restantes que não foram utilizados, eram segmentos muito curtos (de uma unidade lexical ou sem CL), incompletos, repetidos ou que não apresentavam erros de tradução de CL.

Finalmente, fizemos uma análise do processo de anotação da empresa, comparando a nossa anotação de erros de CLs com a anotação feita pelo anotador para verificar similaridades, se os erros anotados pelo anotador envolvem as CLs e de que forma estes erros foram anotados, e fizemos uma análise dos erros de tradução de CLs através da identificação do domínio, descrição e explicação dos erros, a fim de compreender a origem destes erros, possíveis interferências da LP ou LC, etc.

Passemos à descrição do glossário.

### 4.3 Criação de glossário

Os glossários são ótimas ferramentas para a TA, sobretudo no caso das CLs, que são muito desafiadoras para a tradução, e sobretudo para a NMT, que consiste na autoaprendizagem com base nas informações adicionadas: “*Glossaries are food for our NMT system*” (Unbabel, 2018b).

A fim de contribuir para a correção dos erros de tradução de CLs do PB, foi criado um glossário com a tradução das expressões que constituíram frequentemente problemas no *corpus* analisado. Este glossário contém a tradução de CLs (da língua comum e especializada) de inglês para PB, incluindo casos com diferenças entre PB e PE e apontando a tradução para cada variedade.

Este capítulo tratou da metodologia utilizada na pesquisa, apresentando o processo de anotação da empresa, o sistema “*Annotate*”, a métrica de qualidade utilizada e os critérios de anotação. Também foi feita a descrição do *corpus* analisado e do glossário criado. Para o estudo da tradução das CLs, a análise do processo de anotação é muito importante, visto que, no momento em que esta pesquisa foi feita, não havia um tratamento específico para a anotação destas expressões. Baseado neste processo de anotação, o próximo capítulo trata da análise dos dados.

## 5 Análise dos dados

A análise dos dados está dividida em duas partes. A primeira parte diz respeito ao processo de anotação. Fizemos uma anotação dos erros de tradução das CLs e a comparamos com a anotação de erros feita pelo anotador da empresa para os três tipos de erro, a fim de verificar se os erros anotados pelo anotador estão relacionados com as CLs e, caso positivo, como este procedeu com a anotação destas expressões.

A segunda parte, inspirada pelo modelo de Alonso Ramos *et al.* (2010), também utilizado por Costa (2017), trata da análise dos erros de tradução das CLs encontradas no *corpus*, do ponto de vista linguístico, com a finalidade de encontrar a origem destes erros. Assim, foi feita a descrição dos erros e foi dada uma possível explicação para os mesmos.

Essencial para ambas as análises também foi a identificação dos domínios, no caso das CLEs. Assim, o capítulo está dividido da seguinte forma: primeiramente, na seção 5.1 é feita a identificação dos domínios; em seguida, a seção 5.2 apresenta a análise do processo de anotação; a seção 5.3 apresenta a análise dos erros de tradução; e por último, com base nos resultados deste trabalho, a seção 5.4 apresenta as sugestões de implementação de melhorias para a empresa.

Ressaltamos, mais uma vez, que esta análise foi feita integralmente com base no *corpus* recebido, sem acesso a quaisquer instruções ou glossário de clientes a que, por vezes, os anotadores têm acesso. Em alguns casos, apenas os dados recebidos podem não ser suficientes para interpretar os segmentos de texto corretamente, pois há apenas um trecho do e-mail ou da comunicação que constitui os TP e TC. Por exemplo, no caso de “*shop chests*” traduzido por “cofres da loja” no domínio dos jogos, não é possível saber se esta tradução está incorreta ou se está de acordo com alguma instrução ou glossário do cliente, o que pode ter sido uma limitação.

## 5.1 Identificação dos domínios

No seção 3.3.2, referimos que a identificação dos domínios das CLEs é muito relevante para a TA, visto que pode contribuir para a desambiguação, já que determinada palavra tende a combinar com certas palavras em um domínio e com outras palavras em outro domínio, facilitando a distinção desses domínios.

No *corpus* analisado, notamos que a falta da indicação do domínio das CLs é realmente problemática para a TA e para a pós-edição, sobretudo, nos casos de ambiguidade, conforme os exemplos da tabela 10:

Tabela 10  
*Exemplo de ambiguidade*

N ° do exemplo	Tipo de erro	Nível	Erro anotado	LP (inglês)	LC (PB)	Sugestão de tradução das CLs (PB)
1	Lexical Selection	major	bilhete	If there is nothing else let me know and I'll mark this <b>ticket</b> as solved.	Se não houver nada mais, avise-me e marcarei este [ <b>bilhete</b> ] como resolvido	tíquete (de suporte) / ticket
2	Lexical Selection	major	arquivar	If you notice an unauthorised transaction on your account please try to contact the merchant first and ask for the refund as we cannot cancel any card payments, If you are not able to contact the merchant or they refuse to refund you the payment, come back to us and we can <b>file a chargeback request</b> as a last resort.	Se você notar uma transação não autorizada na sua conta, primeiro entre em contato com o vendedor e peça o reembolso, pois não podemos cancelar os pagamentos em cartão; Se você não conseguir entrar em contato com o vendedor ou se ele se recusar a reembolsar o pagamento, fale conosco novamente e nós podemos [ <b>arquivar</b> ] <b>um pedido de estorno</b> como último recurso.	apresentar / abrir um pedido de estorno

No exemplo 1, “*ticket*” (de suporte) da área de TI foi traduzido como “bilhete”, passando do domínio de TI para o domínio dos transportes, por exemplo.

No exemplo 2, a palavra “*file*” foi traduzida como “arquivar”. Na área jurídica, o verbo “*file*” é equivalente a “arquivar” em PB, por exemplo, “*file a lawsuit*” e “arquivar um processo”. No entanto, o TP pertence à área financeira, e deste modo, a palavra “arquivar” altera completamente o sentido. No dicionário “*Cambridge*”, a palavra “*file*” inclui a seguinte definição, relativamente à contabilidade: “*to officially send something such as your accounts: Any gain or loss should be reported when you **file** your tax return*”. No dicionário “*Michaelis*”, os significados de “arquivar” são apenas: “depositar ou guardar em arquivo”, “conservar ou reter na memória”, na área jurídica: “encerrar o andamento de (processo, inquérito, etc) e na área da informática: “copiar arquivos para que fiquem armazenados pelo tempo que se desejar”. Ou seja, os significados da palavra na LC são justamente o oposto do significado existente no TP (que é da área financeira) e não incluem uma definição equivalente nesta área.

Apesar da identificação dos domínios contribuir para a desambiguação, também referimos que apenas a identificação dos domínios não facilita o processo de escolhas de tradução das CLEs, pois estas expressões também envolvem a restrição combinatória e são determinadas pelo uso (cf. L’Homme, 2000; Valente, 2005).

Assim, nesta seção, fizemos a identificação dos domínios, contribuindo para a desambiguação, mas, nesta pesquisa, também levamos em conta a questão da restrição combinatória e é com esta noção que as seções seguintes analisam os dados.

A figura 8 mostra as áreas em que os erros de CLs do *corpus* foram encontrados.



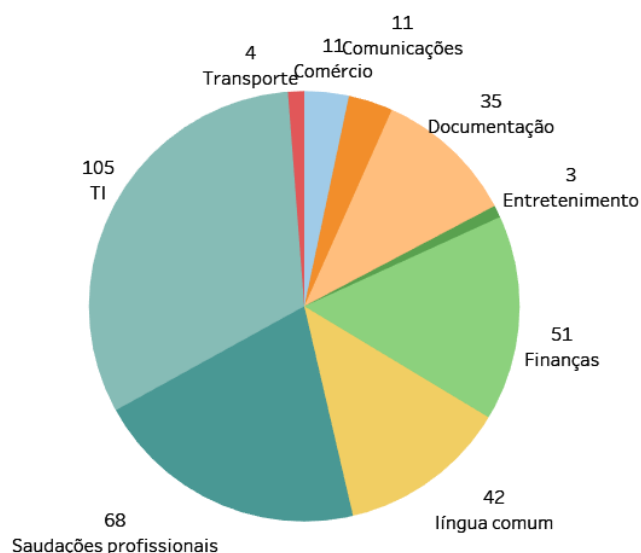


Figura 8: Domínios

A maioria dos erros de CLs são da área de TI, 105 erros. Conforme referido na seção 3.3.3.4, esta área apresenta muitos problemas devido à crescente inclusão de novos termos e expressões e às diferenças entre as expressões equivalentes em PB e PE.

Em segundo lugar, foram observados bastantes erros nas saudações profissionais, (68), que são as expressões utilizadas frequentemente como saudações nos e-mails corporativos, normalmente usadas no registro formal. Por exemplo, as fórmulas de cumprimento “Prezado”, “Estimado” ou “Caro”, as fórmulas de despedida “Cordialmente”, “Atenciosamente” e as fórmulas de pré-despedida, como: “Agradeço mais uma vez a sua compreensão”. As saudações, de forma geral, são expressões muito idiossincráticas e determinadas pelo uso. Por exemplo, em português dizemos “boa noite” no singular, já em espanhol “*buenas noches*” usa-se no plural.

Em seguida, foram encontrados mais erros na área das finanças, 51 erros. A tabela 11 mostra alguns exemplos.

Tabela 11  
Finanças

N ° do exemplo	LP (inglês)	LC (PB)	Sugestão de tradução da CL (PB)
3	account team	equipe da nossa conta	equipe de contas
4	Please forward the following details to allow us to transfer this money to you or your <b>nominated account</b> :	Por favor envie os seguintes detalhes para nos permitir transferir esse dinheiro para você ou sua <b>conta [nomeada]</b> :	conta vinculada / adicionada / cadastrada

No exemplo 3, a tradução muda completamente o sentido da expressão no TP e no exemplo 4, a expressão foi traduzida literalmente.

Em quarto lugar, estão as CLs não especializadas. Foram encontrados 42 erros da língua comum. Depois, 35 erros foram encontrados em relação à documentação. A tabela 12 mostra alguns exemplos da área de documentação.

Tabela 12  
Documentação

N ° do exemplo	LP (inglês)	LC (PB)	Sugestão de tradução da CL (PB)
5	driving licence	carta de condução	Carteira Nacional de Habilitação (CNH), Carteira / carta de motorista, Carteira / carta de habilitação
6	proof of address	prova de endereço	comprovante de endereço

O exemplo 5 envolve erro de variedade linguística pois “carta de condução” é a expressão usada em PE. No PB, o nome do documento é “Carteira Nacional de Habilitação (CNH)”, também conhecida por “carteira / carta de motorista ou de habilitação”. O exemplo 6 é outro caso de tradução literal.

As áreas de comunicações e comércio apresentam a mesma quantidade de erros, 11 ocorrências em cada (tabela 13).

Tabela 13  
*Comunicações e Comércio*

N ° do exemplo	LP (inglês)	LC (PB)	Sugestão de tradução da CL (PB)
7	mobile operator	operador móvel	operadora de telefonia móvel
8	return department	departamento de retorno	departamento de devoluções

O exemplo 7 ilustra um erro da área das Comunicações e o exemplo 8, do Comércio. Em ambos os casos, a tradução foi literal.

Por fim, 4 erros foram encontrados na área de Transporte e foram encontrados 3 erros na área de Entretenimento (filmes e jogos, no caso do *corpus* analisado). A tabela 14 mostra exemplos destas duas áreas, respectivamente.

Tabela 14  
*Transporte e Entretenimento*

N ° do exemplo	LP (inglês)	LC (PB)	Sugestão de tradução da CL (PB)
9	estimated arrival	chegada estimada	previsão de chegada
10	based on chance	com base no acaso	de modo aleatório

No exemplo 9, novamente, a tradução foi literal e no exemplo 10, havia uma opção mais apropriada para o contexto dos jogos.

Identificados os domínios, a seção a seguir é dedicada à análise das CLs no processo anotação.

## 5.2 A anotação dos erros

A análise da anotação dos erros do *corpus* foi feita a fim de verificar a frequência de erros de tradução de CLs, a maneira como estes erros foram anotados e os problemas envolvidos neste processo.

Conforme referido no início do capítulo, fizemos inicialmente uma anotação dos erros relativamente às CLs para então compará-la aos erros anotados no *corpus* recebido e verificar como estas expressões foram anotadas pelo profissional responsável.

### 5.2.1 Frequência dos erros de combinatórias lexicais

Ao compararmos o total de segmentos anotados pelo anotador para os três tipos de erro com o total de segmentos de erros de CLs que anotamos, apuramos a quantidade de segmentos de texto com erros de CLs no *corpus* e analisamos a distribuição desses erros entre os tipos observados. A tabela 15 ilustra a quantidade de erros de tradução das CLs em cada tipo de erro analisado.

Tabela 15  
*Quantidade de erros de tradução de combinatórias lexicais por tipo de erro*

<b>Tipo de erro</b>	<b>Total geral de segmentos</b>	<b>Total de segmentos com erros de CLs</b>	<b>Percentual de erros de CLs por tipo de erro</b>
<i>Lexical Selection</i>	282	187	66,31%
<i>Overly Literal</i>	127	81	63,77%
<i>Wrong Language Variety</i>	65	52	82,53%
Total	474	320	67,51%

Relativamente à frequência, foram encontrados erros de tradução de CLs na maior parte do *corpus* analisado, em 67,51% dos segmentos. Ou seja, dos 474 segmentos de erros anotados

dentro dos três tipos analisados (“*Lexical Selection*”, “*Overly Literal*” e “*Wrong Language Variety*”), 320 segmentos apresentam erros de tradução de CLs.

Conforme referido no capítulo 3, de acordo com Monti e Todirascu (2015), é comum os sistemas de TA traduzirem as unidades multi-palavras incorretamente (seja qual for a abordagem: estatística, baseada em regras, etc.). E o desafio não é só para os sistemas de TA, mas também para o processo de anotação dos erros (manual ou não). Isto ocorre por diferentes razões, como, por exemplo, a possível distância colocacional, a falta de uma tipologia de erros adequada (e critérios de anotação), as propriedades complexas das CLs, o fato de serem diferentes nas diferentes línguas, etc.

Em relação à distribuição dos erros, cada tipo de erro também possui a maior parte dos segmentos com erros de tradução de CLs, sendo a maior porcentagem de erros no tipo “*Wrong Language Variety*”, 82,53% dos segmentos. Os tipos de erro “*Lexical Selection*” e “*Overly Literal*” apresentam uma porcentagem semelhante de segmentos com erros de CL, 66,31% na primeira e 63,77% na última. Adicionalmente, pode haver variação nestes resultados quanto à repetição dos erros, pois o *corpus* recebido reflete os tipos de erros mais frequentes, ou seja, que se repetem mais vezes, e dentro destes tipos há também casos de erros repetidos. No entanto, a análise da frequência de cada erro presente no *corpus* não foi feita, pois não era o foco da pesquisa. De qualquer forma, na análise da origem dos erros (disponível na seção 5.3) foi calculada a frequência por tipo de erro de CL.

### **5.2.2 Anotação dos erros de combinatórias lexicais**

Após analisar a quantidade de segmentos com erros de CLs e a sua relação em cada tipo de erro anotado, fizemos uma análise mais detalhada do *corpus* em relação à anotação dos erros de tradução dessas expressões. Verificamos se os erros anotados pelo anotador são erros de CLs e, caso positivo, como estas anotações foram feitas.

Nos segmentos em que encontramos erros de CLs e os anotamos (320 segmentos), descartamos os segmentos repetidos que se referiam ao mesmo erro de CL que anotamos, restando assim 286 segmentos. Nestes segmentos, comparamos a anotação feita pelo anotador com a anotação que fizemos para verificar se eram referentes ao mesmo erro. A tabela 16 mostra os resultados.

Tabela 16

*Comparação entre as anotações do corpus e as anotações de erros de combinatórias lexicais*

### Anotação referente ao mesmo erro?

	Número de segmentos	%
não	41	14.34%
sim	245	85.66%
Total	286	100.00%

Ao comparar a nossa anotação de erros de CLs com a anotação de erros em geral feita pelo anotador, muitos dos erros anotados pelo anotador estão relacionados com os erros de CLs que anotamos, 85,66% dos casos. Ou seja, nestes casos, o erro anotado pelo anotador também foi erro de tradução de CL (245 de 286 segmentos).

Relativamente à maneira como o anotador anotou estes erros de CLs, foi observada uma inconsistência na extensão da anotação. Por vezes, foi anotada apenas uma unidade lexical, outras vezes, mais de uma unidade. Assim como foi notada uma incoerência na escolha dos tipos de erros. A tabela 17 mostra alguns exemplos:

Tabela 17  
Inconsistência na anotação das unidades lexicais

N ° do exemplo	Tipo de erro	Nível	Erro anotado	LP (inglês)	LC (PB)	Sugestão de tradução da CL (PB)
11	Lexical Selection	major	remoção de pagamentos	Do we understand it correctly that you mean <b>withdrawing funds</b> at the ATMs of (BANK NAME)?	Entendemos corretamente que quer dizer a <b>[remoção de pagamentos]</b> em caixas automáticos do (BANK NAME)?	saque / retirada de fundos
12	Lexical Selection	major	remover	If so, please be informed that you may use the (COMPANY) card to <b>withdraw cash</b> or make payments at any ATM / POS-terminal that accepts (CARD BRAND) cards.	Se assim for, por favor, fique ciente de que pode usar o cartão do (COMPANY), de modo a <b>[remover]</b> o dinheiro ou fazer pagamentos em qualquer terminal que aceite pagamentos (CARD BRAND).	sacar / retirar o dinheiro
13	Lexical Selection	major	Lamento ouvir	<b>I'm sorry to hear</b> that your design didn't save.	<b>[Lamento ouvir]</b> que o seu design não foi salvo.	sinto muito
14	Overly Literal	major	ouvir	Thanks for reaching out to us and <b>I'm sorry to hear</b> that your design didn't save.	Obrigado por nos contatar e lamento <b>[ouvir]</b> que o seu design não tenha sido salvo.	

Os exemplos 11 e 12 são CLEs da área financeira e os exemplos 13 e 14 são CLs com função pragmática, muito frequentes nos e-mails profissionais.

No exemplo 11, a CL toda foi anotada como incorreta “remoção de pagamentos”, enquanto que no exemplo 12, apenas a palavra “remover” foi anotada. Ambos os exemplos foram anotados como “*Lexical Selection*”, nível “*major*”.

Em relação ao tipo de erro selecionado, já referimos que, de acordo com as *Guidelines*, o tipo de erro “*Lexical Selection*” deve ser utilizado quando o termo selecionado não está adequado para determinado contexto ou não transmite precisamente o significado do texto. Talvez este tipo de erro não seja o mais apropriado para a anotação das CLs e anotar essas expressões conforme um tipo de erro específico para tais seja uma melhor opção. Pois, embora

as CLEs possam ser compostas por termos complexos, como já referido, o mesmo não ocorre nas CLs da língua comum, como é o caso dos exemplos 13 e 14, de modo que esta descrição não se aplica a estas combinatórias. Como referimos atrás, Costa *et al.* (2015) propõem uma nova taxonomia de erros linguisticamente motivada que inclui, especificamente, os tipos “*collocational\_errors*” e “*Idioms*”. Estes tipos de erros selecionam os erros exclusivamente em blocos de palavras, o que pode ser uma boa solução para a anotação destas expressões. Em “*collocation errors*”, os autores dão o exemplo da anotação desse tipo: “*high wind*” em inglês, traduzido por “vento alto” e a tradução correta “vento forte”.

Isto porque, quanto à forma de anotação destes exemplos, quando apenas uma parte da CL é anotada, é preciso ler o texto para entender o contexto e a razão pela qual a palavra anotada está incorreta naquele caso, enquanto que apontar a CL toda como incorreta é autoexplicativo. Isto se aplica não só ao contexto da anotação; de acordo com Elloumi *et al.* (2015), considerar as CLs como um todo força a segmentação durante o alinhamento, contribuindo, assim, para a detecção automática de CLs e para a melhoria dos sistemas de TA.

Mesmo no caso das colocações que, diferentemente das EIs, por possuírem um certo grau de variabilidade, de acordo com alguns autores, não formam uma unidade semântica (como citamos Aisenstadt (1979 *apud* Fontenelle, 1994) no capítulo 3), elas são arbitrárias, ao contrário das combinações livres de palavras, seus elementos não são livremente comutáveis e por isso não devem ser traduzidas (e anotadas) palavra a palavra. Assim, “*Collocational errors occur in blocks of words*” (Costa *et al.*, 2015, p. 137).

Além disso, segundo a definição de “erro” nas *Guidelines*, “*An error represents any issue that you may find in the translated text that either does not correspond to the source text or is considered incorrect (i.e. agrammatical) in the target language*” (Unbabel, 2018b, p. 8). Neste sentido, no exemplo 12, não é a palavra “remover” que está incorreta na LC, visto que é



uma palavra existente na LC, e sim a CL “remover o dinheiro” está incorreta, pois esta não é a forma utilizada nesta língua e sim “sacar / retirar o dinheiro”.

A mesma inconsistência ocorre nos exemplos 13 e 14, com a anotação da CL “lamento ouvir” no exemplo 13 e apenas da palavra “ouvir” no exemplo 14. Apesar disto, a forma mais frequente de anotação observada foi a de apenas uma unidade lexical, como veremos mais adiante.

Nos exemplos 13 e 14, notamos também uma inconsistência na escolha do tipo de erro, “*Lexical Selection*” foi utilizada para o exemplo 13 e “*Overly Literal*” para o exemplo 14. Conforme mencionado no capítulo 3, estes dois tipos de erro podem ser confusos para o anotador, já que os seus usos não estão muito claros nas *Guidelines*. Apesar disso, segundo os critérios, o tipo “*Overly Literal*” deve ser utilizado quando a tradução fica muito restrita ao TP, podendo causar problemas de interpretação (como a tradução literal de EI). E a regra para a tradução de expressões é que, se uma parte da expressão foi traduzida muito literalmente, a expressão toda deve ser anotada. Ou seja, a anotação desse tipo de erro deve ser da expressão toda e não de apenas uma palavra, como foi feito no exemplo 14. Outros exemplos são observados na tabela 18.

Tabela 18  
Inconsistência na anotação de expressões

N ° do exemplo	Tipo de erro	Nível	Erro anotado	LP (inglês)	LC (PB)	Sugestão de tradução da CL (PB)
15	Lexical Selection	major	Estou ansioso	I look forward to hearing back from you!	[Estou ansioso] para ouvir de você!	Aguardo retorno.
16	Lexical Selection	major	para ouvir	I look forward to hearing back from you!	Estou ansioso [para ouvir] de você!	
17	Lexical Selection	minor	Espero	Looking forward to your reply,	[Espero] sua resposta,	
18	Overly Literal	major	Espero por sua resposta	Looking forward to your reply,	[Espero por sua resposta],	

Nos exemplos 15 e 16, a expressão “*I look forward to hearing back from you*”, traduzida por “Estou ansioso para ouvir de você”, foi anotada duas vezes diferentes (ambas como tipo “*Lexical Selection*”, nível “*major*”), uma anotação de “estou ansioso” e outra de “para ouvir”, quando, na verdade, a expressão toda deveria ter sido anotada uma única vez, visto que se trata de uma CL com função pragmática, ou seja, convencionada a determinada situação (neste caso, característica do encerramento de e-mails corporativos) e, por isso, é uma expressão fixa. Como refere Costa (2017), a fixidez abrange este tipo de CL.

O mesmo tipo de CL está presente no caso dos exemplos 17 e 18 e a tradução da expressão “*Looking forward to your reply*,” também foi anotada duas vezes. Em 17, apenas a palavra “Espero” foi anotada, como “*Lexical Selection*” e nível “*minor*”, enquanto que em 18, a expressão toda foi anotada e classificada como “*Overly Literal*”, nível “*major*”.

A inconsistência na escolha do tipo de erro também foi notada em relação a “*Wrong Language Variety*”, como mostra a tabela 19:

Tabela 19  
Inconsistência na anotação dos tipos de erro

N ° do exemplo	Tipo de erro	Nível	Erro anotado	LP (inglês)	LC (PB)	Sugestão de tradução da CL (PB)
19	Wrong Language Variety	major	aplicação	I have <b>closed this application</b> as you seem to be stuck in sign up process.	Eu <b>fechei esta [aplicação]</b> como você parece estar preso no processo de inscrição.	fechei o aplicativo
20	Lexical Selection	major	aplicação	I have <b>closed this application</b> as you seem to be stuck in sign up process.	Como parece que você está preso no processo de inscrição, eu <b>fechei essa [aplicação]</b> .	

Ambos os erros são de variedade linguística, pois o termo “aplicação” da área de TI é usado em PE e equivale, em PB, a “aplicativo”. Contudo, no exemplo 19, o erro foi anotado como “*Wrong Language Variety*” e no exemplo 20, o mesmo erro é anotado como “*Lexical Selection*”. Aliás, muitos erros de “*Wrong Language Variety*” foram, na verdade, anotados como “*Lexical Selection*” ou mesmo “*Overly Literal*”. (Na seção 5.3.2, veremos que os erros de variedade linguística representam aproximadamente um terço dos erros de CLs). Uma possível razão para isto acontecer pode ser o conhecimento insuficiente de outras variedades linguísticas por parte do anotador para poder identificar que a origem do erro está na variedade utilizada. Já referimos, no capítulo 3, o exemplo de Costa (2017) em que casos de usos do espanhol da América latina foram inicialmente anotados como incorretos e depois como corretos em determinadas variedades linguísticas, quando foram identificados em outros países de língua espanhola.

Outra inconsistência encontrada foi em relação aos segmentos em que havia mais de um tipo de erro. Nestes casos, as *Guidelines* não esclarecem como deve ser feita a anotação (há instrução para os casos em que o erro se enquadra em dois tipos diferentes, neste caso, a orientação é para que seja anotado o erro o mais grave). No momento em que este estudo foi

feito, não era possível anotar mais de um tipo de erro por segmento de texto. De qualquer maneira, observamos que, por vezes, foram feitas diferentes anotações para o mesmo segmento de texto (vide tabela 20).

Tabela 20  
*Segmentos com mais de uma anotação*

Quantos segmentos foram anotados mais de uma vez?

	Número de segmentos
anotados duas vezes	31
anotados três vezes	5
anotados quatro vezes	1
anotados cinco vezes	1
Total	38

Ou seja, 38 segmentos foram anotados mais de uma vez. Um segmento aparece anotado cinco vezes diferentes, uma para cada erro. Outro segmento foi anotado quatro vezes. 5 segmentos foram anotados três vezes diferentes e 31 segmentos foram anotados duas vezes diferentes (incluindo os já mencionados exemplos 15 e 16, de erros de CLs, em que poderia ter sido feita uma única anotação em vez de duas).

Há casos em que, ao invés de se fazer mais de uma anotação para o segmento, foi feita a anotação do erro mais grave e em outros casos, não foi o erro mais grave que foi anotado. Os casos em que o erro mais grave foi anotado são compatíveis com os erros que anotamos de CLs. Ou seja, nestes casos, os anotadores consideraram os erros relacionados com as CLs os mais graves dos segmentos anotados, o que reforça a relevância deste tipo de erros.

No entanto, como mencionado anteriormente, muitos destes erros foram anotados apenas parcialmente pelo anotador, como mostra a tabela 21.

Tabela 21  
*Como os erros de combinações lexicais foram anotados*

	Número de segmentos	%
Anotação demasiada	8	3.27%
Mais de uma palavra, mas anotação parcial	29	11.84%
Anotação da combinação lexical toda	80	32.65%
Uma palavra anotada	128	52.24%
Total	245	100.00%

Na maior parte dos casos, a anotação foi apenas de uma unidade lexical (52%). Ou seja, dos 245 segmentos anotados que coincidem com a nossa anotação de CLs, 128 foram anotados por uma única palavra. Como identificamos que são erros de CLs, a anotação parcial não é a mais adequada para esse tipo de erros. Apenas, em 32% dos casos a anotação foi da CL toda, procedimento adequado para estas expressões. Em 11% dos casos mais de uma palavra foi anotada, mas a anotação foi ainda parcial, e, por último, em 3% a anotação foi além do necessário.

Nos erros de CLs em que apenas uma unidade lexical foi anotada, notamos também casos em que havia distância colocacional, conforme mencionado no capítulo 3, este é um dos desafios para a TA. De acordo com as *Guidelines*, “*if there are two or more units spread apart within the text that form a single error, then you will need to select all the units and choose one type of error*” (Unbabel, 2018b, p. 12). Isto pode ser aplicado, então, aos casos em que há distância colocacional, em vez de anotar apenas uma unidade lexical. No *corpus* analisado, não foi encontrado nenhum erro anotado levando em conta a distância colocacional. A tabela 22 mostra alguns exemplos:

Tabela 22  
Distância colocacional

N ° do exemplo	Tipo de erro	Nível	Erro anotado	LP (inglês)	LC (PB)	Sugestão de tradução da CL (PB)
21	Overly Literal	major	retornar	The <b>items</b> you wish to <b>return</b> must be sent back within 30 days of receiving them.	Os <u>itens</u> que você deseja <b>[retornar]</b> devem ser enviados no prazo de 30 dias após o recebimento.	Os <u>itens</u> que você deseja <b>devolver</b> ...
22	Overly Literal	major	aplicação	Your <b>application</b> to create an account is under the <b>review</b> .	A sua <b>[aplicação]</b> para criar uma conta está sendo <b>revista</b> .	O seu <b>aplicativo</b> para criar uma conta está sendo <b>analisado</b> .
23	Lexical Selection	critical	levantado	Keep in mind that the <b>shipment</b> can not be cancelled if it <b>has been picked up</b> .	Tenha em mente que o <b>frete</b> não pode ser cancelado se já tiver sido <b>[levantado]</b> .	Observe que o <b>frete</b> não pode ser cancelado se já tiver sido <b>retirado</b> .

Estes três exemplos pertencem às CLEs. O exemplo 21 é da área comercial, o exemplo 22 de TI e o exemplo 23 é da área de transportes.

No exemplo 21, apenas a palavra “retornar” foi anotada, no exemplo 22, apenas a palavra “aplicação” e no exemplo 23, apenas a palavra “levantado”. Nos exemplos 21 e 23, de fato, são apenas estas palavras que foram afetadas pelo erro, porém, como já referido, a interpretação das CLs ocorre de forma integral, da combinatória como um todo, portanto, talvez a anotação como tal seja a opção mais adequada. Uma das vantagens do processo de anotação de erros é justamente contribuir para a resolução dos erros e, desta forma, a anotação dos erros de CL feita na CL toda reflete num melhor tratamento destas expressões. No caso do exemplo 22, a CL toda está incorreta de qualquer forma, a tradução neste caso deveria ser: “o aplicativo para criar uma conta está sendo analisado”. De forma que a CL toda deveria ter sido anotada, as palavras “aplicação” e “revista”, que estão separadas por uma distância colocacional.

Além disso, os exemplos 22 e 23 foram anotados como tipo “*Overly Literal*” e “*Lexical Selection*” respectivamente, no entanto, ambos os exemplos envolvem também “*Wrong Language Variety*”, pois, “aplicação” e “levantar o frete” são usados no PE, no PB os

equivalentes são “aplicativo” e “retirar o frete”. Novamente, detectar as variedades linguísticas ainda é um desafio para a TA, sobretudo no caso das CLs, que possuem propriedades tão complexas e, por isso, estendem a dificuldade também aos tradutores / revisores / anotadores humanos que não têm o nível de conhecimento elevado na outra variedade linguística que estas expressões refletem.

Observamos ainda um caso em que a CL foi desmontada e anotada três vezes, uma para cada unidade que a compõe:

Tabela 23  
*Exemplo de mesmo segmento anotado três vezes*

N ° do exemplo	Tipo de erro	Nível	Erro anotado	LP (inglês)	LC (PB)	Sugestão de tradução da CL (PB)
24	Lexical Selection	major	suportamos	We do not <b>support standing orders</b> as of yet but it will be available in few weeks.	Ainda não <b>[suportamos] ordens permanentes</b> , mas estará disponível em algumas semanas.	Ainda não <b>atendemos ordens permanentes</b> , mas este serviço estará disponível em algumas semanas.
25	Lexical Selection	major	ordens	We do not <b>support standing orders</b> as of yet but it will be available in few weeks.	Ainda não <b>suportamos [ordens] permanentes</b> , mas estará disponível em algumas semanas.	
26	Lexical Selection	major	permanentes	We do not <b>support standing orders</b> as of yet but it will be available in few weeks.	Ainda não <b>suportamos ordens [permanentes]</b> , mas estará disponível em algumas semanas.	

Nos exemplos 24, 25 e 26 da tabela 23, cada palavra da CL foi anotada individualmente, em vez de anotar a CL toda uma única vez. “Ordem permanente”, apesar de ter sido anotada como erro, é a expressão da área financeira equivalente ao TP. O que foi traduzido de forma literal indevidamente neste caso foi o verbo “*support*”, que não é expresso desta forma no TC, neste contexto.

Por fim, a tabela 24 mostra outro caso de anotação parcial de CL e falha na detecção da variedade linguística.

Tabela 24

Exemplo de anotação parcial e tipo de erro incorreto

N ° do exemplo	Tipo de erro	Nível	Erro anotado	LP (inglês)	LC (PB)	Sugestão de tradução da CL (PB)
27	Lexical Selection	major	levanta	Please note that when you <b>withdraw cash</b> or make payments not in the card's currency ((CURRENCY)), the additional fee of (PERCENTUAL)% is applied.	Por favor, note que quando [ <b>levanta</b> ] dinheiro ou efetua pagamentos que não sejam na moeda do cartão ((MOEDA)), a taxa adicional de (PERCENTUAL)% é aplicada.	saca / retira dinheiro

No exemplo 27, o erro da CL “levantar dinheiro” foi anotado apenas na unidade lexical “levanta” e como “*Lexical Selection*” quando é, na verdade, erro de variedade linguística, pois conforme referido na seção 3.3.3, esta é a expressão utilizada em PE, que possui um sentido figurado convencionado (levantar). No PB a equivalente é “sacar / retirar dinheiro”.

Estes foram os resultados da análise da anotação do *corpus*, identificamos que os erros de tradução de CL são muito frequentes no *corpus* analisado e que há muita inconsistência na forma de anotação dos mesmos, tanto na extensão da anotação quando na escolha do tipo de erro que representam. Também analisamos os problemas envolvidos na anotação destas expressões e apontamos possíveis soluções, principalmente, a implementação de um tipo de erros específico para as CLs como referimos “*collocational error*” de Costa *et al.* (2015), assim como a criação de critérios de anotação para estas expressões.

A próxima seção apresenta a análise, precisamente, dos erros de tradução de CLs.



### 5.3 Erros de tradução de combinatórias lexicais

Esta análise foi inspirada pelo modelo de Alonso Ramos *et al.* (2010), também utilizado por Costa (2017), a partir do ponto de vista linguístico, que busca identificar a origem do erro. Assim como Costa, analisamos unicamente os erros lexicais e utilizamos, na descrição dos erros, o tipo “opção melhor” que a investigadora acrescentou, bem como, na explicação dos erros, o tipo de erro “variedade linguística” (“variante” na designação da autora).

No entanto, diferentemente da análise de Costa (2017) e Alonso Ramos *et al.* (2010), que abordam apenas as colocações, este trabalho analisa também outros tipos de CLs, pois, como já foi mencionado, os problemas que estas unidades acarretam para a tradução são os mesmos e todas diferenciam-se das combinatórias livres de palavras.

Apesar disso, as colocações foram, de qualquer forma, a maioria dentre as unidades encontradas no *corpus* desta pesquisa devido à transparência semântica e ao fato de que muitas das expressões encontradas no *corpus* são CLEs e, portanto, conforme já referido, as EIs não são comuns neste tipo de combinatórias. Mas, novamente, este trabalho opta por não fazer a distinção das unidades e utilizar apenas a designação “Combinatórias Lexicais”. Por outro lado, também não fizemos a análise “localização das colocações” que as autoras fizeram, pois não era o foco desta pesquisa.

Finalmente, a escolha deste modelo também se dá pelo fato de que este é baseado no ponto de vista linguístico e não computacional como os trabalhos no ramo da TA tendem a apresentar, e permite uma explicitação dos erros nas perspectivas interlíngua e intralíngua, como é o objetivo deste trabalho.

Assim sendo, a seção 5.3.1 apresenta a descrição dos erros e a seção 5.3.2 apresenta as possíveis explicações para a sua ocorrência.

### 5.3.1 Descrição dos erros

No âmbito da TA, a análise dos tipos de erros de CLs pode servir para a criação de ferramentas de correção automática que preveem os possíveis erros de tradução.

Os erros lexicais do modelo referido atrás podem ser erros de **“substituição”**, **“criação”**, **“síntese”**, **“análise”**, **“sentido diferente”**, **“opção melhor”** e **“utilização incorreta de preposição”** (cf. Costa, 2017).

O erro de “substituição” refere-se à troca de palavras por outras que também existem na LC. “Criação” refere-se à utilização de uma palavra que não existe na LC. O erro de “síntese” ocorre quando uma única palavra é utilizada em vez de uma CL. A “análise” corresponde ao inverso da síntese, a utilização de uma expressão quando deveria ser apenas uma palavra. O erro “sentido diferente” consiste na utilização de uma CL correta na LC, mas inadequada para o contexto. Neste último tipo de erro, também consideramos a utilização de uma palavra existente na LC, mas utilizada inadequadamente na CL. No tipo “opção melhor” as CLs não estão propriamente incorretas, mas existe uma opção mais apropriada. Por fim, “utilização incorreta de preposição” ocorre quando uma preposição deveria ter sido utilizada ou foi utilizada indevidamente na expressão.

A figura 9 apresenta a descrição dos erros identificados no *corpus*, na sua totalidade:

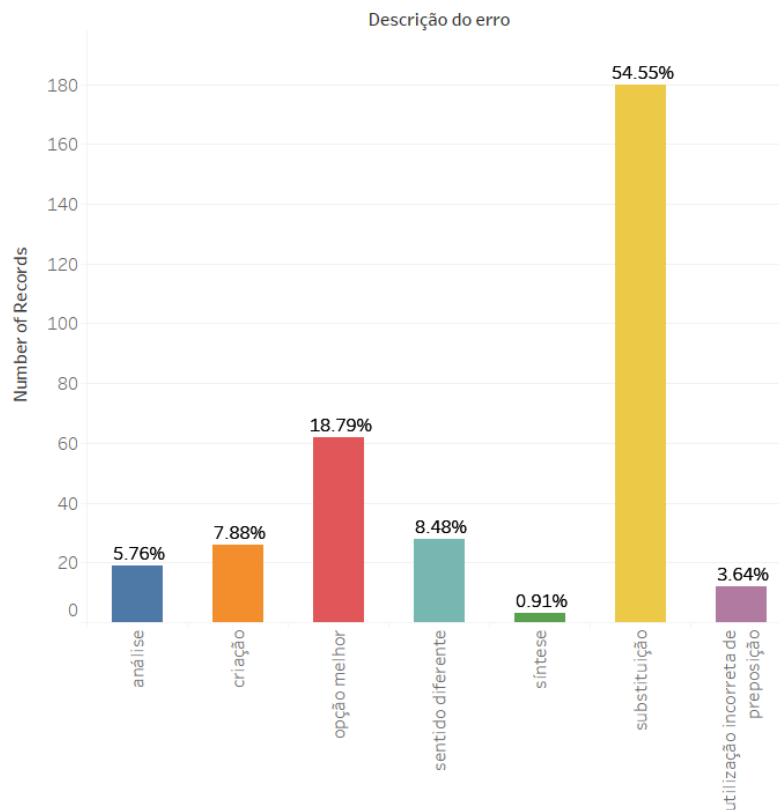


Figura 9: Descrição dos erros

Assim como aconteceu em Costa (2017), os erros mais frequentes são os de substituição, 54%. Trata-se de erros em que ocorre uma substituição de palavras por outras palavras inadequadas para o contexto, mas que existem na LC. A tabela 25 mostra alguns exemplos.

Tabela 25  
Erros de substituição

N ° do exemplo	LP (inglês)	LC (PB)	Sugestão de tradução da CL (PB)
28	screenshots	capturas de [ecrã]	capturas de <u>tela</u>
29	<u>give</u> us a call	<u>envie</u> -nos uma chamada	por favor, entre em contato

Este tipo de erros pode ocorrer por interferência da LP ou da LC, como é o caso do exemplo 28, pois “capturas de ecrã” é a expressão equivalente em PE. (Este é um exemplo em que o conceito só é expresso através de uma CL na LC, já que a LP utiliza uma única unidade lexical).

Porém, também encontramos casos em que não foi notado nenhum tipo interferência, como mostra o exemplo 29, com a tradução de “give” para “enviar”. Neste caso, o TP utiliza um verbo suporte “give”. Como referimos atrás, as construções com este tipo de verbo não podem ser traduzidas literalmente, pois os verbos suportes podem ser diferentes nas diversas línguas. Contudo, este não era o caso neste exemplo, pois, em PB, o verbo suporte é equivalente ao utilizado na LP, pois é possível a expressão “dar uma ligada”. Apesar disso, optamos por sugerir uma expressão mais usual e mais formal para o contexto (e-mails profissionais) “por favor, entre em contato”.

O tipo de erro “opção melhor” é o segundo mais frequente, 18%. Este tipo de erro não é, necessariamente, um erro, mas existe uma opção mais natural que um falante nativo utilizaria ou que é mais adequada ao contexto (cf. tabela 26).

Tabela 26  
*Opção melhor*

N ° do exemplo	LP (inglês)	LC (PB)	Sugestão de tradução da CL (PB)
30	May I ask how exactly can I be of help?	Posso perguntar como posso ajudar?	Como posso ajudar?
31	keep in mind that	tenha em mente que	observe que / lembre-se de que

No exemplo 30, “*may I ask*” foi traduzido para “posso perguntar”. Embora esta opção seja possível na LC, não é a uma opção utilizada com frequência. No exemplo 31, a expressão é utilizada, mas existem opções mais adequadas para o registro do texto.

Em terceiro lugar, “sentido diferente” aparece como tipo de erro mais frequente, como mostra a tabela 27. Este tipo de erro é muito frequente nos casos em que há polissemia.

Tabela 27  
*Sentido diferente*

N ° do exemplo	LP (inglês)	LC (PB)	Sugestão de tradução da CL (PB)
32	you can try heading here	você pode tentar o título Aqui	pode tentar vir aqui
33	bank statement	conta bancária	extrato bancário

No exemplo 32, a tradução da expressão “*you can try heading<sup>12</sup> here*” por “você pode tentar o título Aqui”, pode ter sido devido à polissemia da palavra “*heading*”, que também pode significar “título” (de texto). Neste caso, a solução equivalente na LC é uma combinação livre de palavras e não uma CL como na LP. Como referido anteriormente, este é um dos desafios de tradução destas expressões que o tradutor deve levar em consideração, nem sempre o mesmo conceito é expresso através de uma CL nas diferentes línguas.

Já no exemplo 33, não há uma razão aparente que possa ter ocasionado o sentido errado.

Em 7% das ocorrências, observamos erros de “criação”. No contexto de TA, este tipo de erro ocorre quando uma palavra ou mais da LP é mantida na LC (vide tabela 28).

<sup>12</sup> Expressão “*Head for*” no exemplo apresentado: “*to go or cause to go (towards)*”, in British, “*to direct one's way toward*”, in American. Ex: “*she headed for home*” (Collins Dictionary).

Tabela 28  
Criação

N ° do exemplo	LP (inglês)	LC (PB)	Sugestão de tradução da CL (PB)
34	<b>Drivers Licence</b>	<b>Drives</b>	Carteira Nacional de Habilitação (CNH), Carteira / carta de motorista, Carteira / carta de habilitação
35	<b>Pinning</b> large amounts of <b>unwanted or repetitive stuff</b> , posting unsolicited commercial messages in comments/descriptions, or [...] isn't allowed.	[O <b>pinning</b> d]e grandes quantidades de <b>coisas indesejadas ou repetitivas</b> , postagem mensagens comerciais não solicitadas em comentário/descrições ou [...] não são permitidos.	<b>A</b> <b>marcação</b> de grandes quantidades de <b>coisas indesejadas ou repetitivas</b> , a postagem de mensagens comerciais não solicitadas em comentário/descrições ou [...] não são permitidos.

No exemplo 34, o nome do documento foi mantido em inglês. No exemplo 35, da área de TI, também uma palavra foi mantida em inglês inadequadamente.

Os erros de “análise” representam 5% do *corpus*. São erros em que uma expressão com a estrutura de uma CL é criada indevidamente, quando deveria ser usada uma única palavra. A tabela 29 ilustra alguns exemplos de análise:

Tabela 29  
Análise

N ° do exemplo	LP (inglês)	LC (PB)	Sugestão de tradução das CL (PB)
36	currency	moeda corrente	moeda
37	Now go back to the (COMPANY) app and you can enter your passcode to <b>log in</b> .	Agora volte para o aplicativo (EMPRESA) e você pode inserir sua senha para <b>[fazer] acessar</b> .	acessar

Em relação ao exemplo 37, de acordo com Costa (2017), é comum alunos criarem expressões que assumem a configuração de uma colocação com os verbos “fazer” e “tomar”. “Tratando-se de verbos que não têm um valor lexical muito marcado, tornam-se os “alvos” ideais para serem utilizados quando os alunos não têm a certeza qual o colocativo correto que

a base deve selecionar” (Costa, 2017, p. 170). E em sua análise comparativa com os erros de TA, o uso problemático “fazer e tomar” é uma das semelhanças.

A “utilização incorreta de preposição” aparece em 3% dos casos. Este tipo de erro ocorre quando uma preposição é adicionada, utilizada incorretamente ou ausente na CL, por exemplo:

Tabela 30  
*Preposição incorreta*

N ° do exemplo	LP (inglês)	LC (PB)	Sugestão de tradução da CL (PB)
38	tap share	tocar compartilhar	tocar <u>em</u> compartilhar
39	Fill in your <b>email address</b>	Preencha o seu [e-mail <b>endereço</b> ]	endereço <u>de</u> e-mail

O uso das preposições nas CLs pode variar de língua para língua, e isto é também um problema para a TA. Nestes exemplos, as CLEs de TI em inglês não incluem preposições, já no caso da LC (o português) estas expressões utilizam preposições. De maneira que, no exemplo 38, faltou a preposição “em” e no exemplo 39, a ordem das palavras estava incorreta e a preposição “de” estava ausente.

Finalmente, quase um por cento das ocorrências apresentam erros de “síntese”. Estes erros são o oposto dos erros de análise, uma única palavra é utilizada quando deveria ser usada uma CL. Tal é o caso do exemplo da tabela 31:

Tabela 31  
*Síntese*

N ° do exemplo	LP (inglês)	LC (PB)	Sugestão de tradução da CL (PB)
40	scroll down	desloque-se	role para baixo

No exemplo 40, “*scroll down*”, que é a expressão utilizada no contexto de TI, foi traduzida para “desloque-se”, não apenas suprimindo a CL, mas também alterando o sentido da expressão.

### 5.3.2 Explicação dos erros

A fim de apresentar uma análise dos erros mais detalhada, consideramos pertinente não apenas descrever os erros, mas explicitar as fontes dos mesmos.

Conforme já referido, Alonso Ramos (2010) sugere uma classificação de tipos de erros, também utilizada por Costa (2017), estabelecida com base no aspecto que está na sua origem. Esta classificação é adotada neste trabalho, dado que permite dar conta das causas dos erros identificados relacionados com as CLs. Esta classificação de erros divide-se em dois tipos de erro: “interlíngua” (transferência da LP) e “intralíngua” (transferência da LC).

De acordo com a autora, os tipos de erro interlíngua são aqui os erros de “**importação**” e “**extensão**”. Os erros de “importação” ocorrem quando uma unidade lexical é criada na LC partindo de uma unidade lexical da LP. Esta importação pode ser: por adaptação à LC - Costa (2017) dá o exemplo de importação com adaptação de italiano (Erro: “restos romanos” (*resti romani*) Correção: “vestígios romanos”) - ou importação direta (cópia direta da LP). Ressaltamos que ambos foram classificados no *corpus* apenas como importação. Nos erros de “extensão”, o significado de uma unidade lexical existente na LC é estendido. Em alguns casos de extensão, a unidade lexical escolhida na LC é uma tradução válida dessa mesma palavra na LP, porém, não é a opção correta naquele contexto específico.

Os tipos de erro intralíngua são os seguintes: “**escolha errada**”, “**variedade linguística**”, “**sobregeneralização**” e “**derivação errada**”.

O tipo “escolha errada” é utilizado quando não há uma razão aparente ou não houve uma interferência da LP para o erro produzido. O tipo “variedade linguística”, ocorre quando



é utilizada uma expressão de uma variedade da LC. Segundo Costa, este tipo pode, de certa maneira, ser considerado um erro de interlândia, visto que espelha uma transferência de uma variedade da língua materna (outra variedade linguística mesmo, no caso da TA, ou da língua materna do revisor humano, por exemplo). A “sobregeneralização” ocorre quando é utilizada uma unidade mais vaga ou genérica; no caso da TA, podem ser, por exemplo, casos em que a tradução foi feita literalmente. Por fim, “derivação errada” ocorre quando uma expressão inexistente é produzida na LC (seja por analogia com outra forma da LC ou uma adaptação ortográfica da LP). Este tipo de erros é mais comum nas produções humanas, como no caso de aprendentes de uma língua estrangeira.

A figura 10 apresenta a totalidade das possíveis explicações dos erros de CLs encontrados no *corpus*, usando a classificação proposta por Alonso Ramos (2010) e adaptada por Costa (2017).

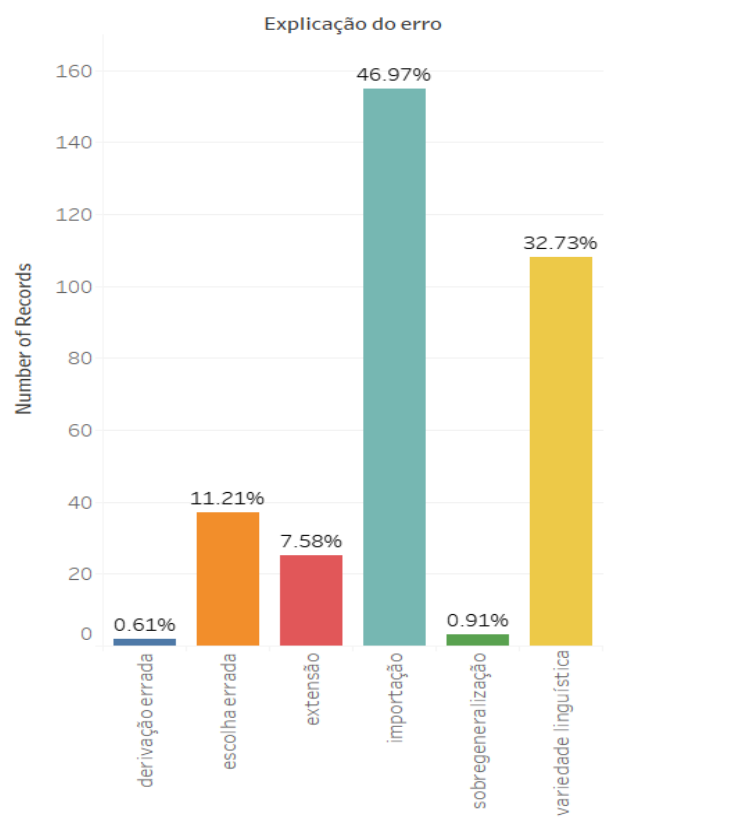


Figura 10: Explicação dos erros

Os erros de “importação” são os mais frequentes (47%). Trata-se, como dissemos atrás, de erros ocasionados por uma interferência da LP. A tabela 32 mostra alguns exemplos de importação:

Tabela 32  
*Importação*

N ° do exemplo	LP (inglês)	LC (PB)	Sugestão de tradução da CL (PB)
41	<u>return</u> department	departamento de <u>retorno</u>	departamento de <u>devoluções</u>
42	When you call please make sure you have your (COMPANY) <b>ID</b> [...]	Quando você ligar, certifique-se de ter seu <b>ID</b> [...]	documento de identificação

No exemplo 41, a importação foi adaptada à LC e no exemplo 42, a importação foi direta, pois “ID” foi mantido em inglês indevidamente na LC.

O segundo tipo de erro mais frequente é “variedade linguística” (33%). Os erros de variedade linguística são os já mencionados casos em que há interferência na LC, como por exemplo, o uso do PE ao invés do PB e vice-versa.

Tabela 33  
*Variedade linguística*

N ° do exemplo	LP (inglês)	LC (PB)	Sugestão de tradução da CL (PB)
43	Looking forward to hear from you soon.	[Fico a aguardar] notícias suas.	Fico no aguardo / aguardo retorno
44	We will be glad to help you.	Teremos [todo o gosto] em ajudá-lo.	Será um prazer ajudá-lo.

Os erros de variedade linguística aparecem tanto em CLs da língua comum quanto em CLEs. É o caso do já referido exemplo 27, em que ocorre a CLE da área financeira “levantar dinheiro” em PE e “sacar dinheiro” em PB e dos exemplos 43 e 44, em que ocorrem CLs características dos e-mails corporativos, cuja fixidez tem uma dimensão pragmática. No

exemplo 43, a utilização do verbo no infinitivo precedido pela preposição *a* é a forma utilizada no PE. Assim como no exemplo 44, a expressão “ter todo o gosto em...”, não é comum no PB.

Os erros por “escolha errada” representam 11% das ocorrências. Este erro ocorre quando não foi encontrada nenhuma interferência da LP ou de outra língua.

Tabela 34  
*Escolha errada*

N ° do exemplo	LP (inglês)	LC (PB)	Sugestão de tradução da CL (PB)
45	give us a call	enviar-nos uma chamada	entre em contato
46	flight data	dados de passagem aérea	dados de voo

No exemplo 45, como já referido na seção anterior, foi feita uma escolha errada do verbo suporte. Também neste caso, o verbo foi deixado no infinitivo indevidamente. No exemplo 46, não há uma razão para a tradução “dados de passagem aérea”, uma vez que a tradução literal, nesse caso, é mesmo “dados de voo”. Além disso, “dados de passagem aérea” muda o sentido do TP, pois pode se referir aos dados da passagem área literalmente.

Em 8% dos casos, foram encontrados erros de extensão. São erros que estendem o sentido de uma palavra da LP, como mostra a tabela 35:

Tabela 35  
*Extensão*

N ° do exemplo	LP (inglês)	LC (PB)	Sugestão de tradução das CL (PB)
47	to log in	para <b>fazer acessar</b>	acessar
48	You can <b>login</b> to your (COMPANY) App	Você pode [ <b>fazer entrar</b> ] no seu aplicativo	acessar / entrar / fazer login

Este tipo de erro é muito frequente na produção das CLs, por conta da já mencionada generalização dos verbos tomar e fazer.

Os erros de “sobregeneralização” representam menos de um por cento. Trata-se de erros nos quais é utilizada uma palavra mais vaga, como o exemplo da tabela 36:

Tabela 36  
*Sobregeneralização*

N ° do exemplo	LP (inglês)	LC (PB)	Sugestão de tradução da CL (PB)
49	To us, spam is any content that's fraudulent or deceptive, <b>pinning</b> large amounts of <b>unwanted or repetitive stuff</b>	Para nós, o spam é qualquer conteúdo fraudulento ou enganador, que <b>[coloque]</b> grandes quantidades de <b>material indesejável ou repetitivo</b>	Para nós, o spam é qualquer conteúdo fraudulento ou enganador, que <b>fixe</b> grandes quantidades de <b>material indesejável ou repetitivo</b>

No exemplo 49, o verbo “colocar” (estabelecer) é mais vago que do “fixar” (estabelecer definitivamente). Além disso, este exemplo trata de uma CLE de TI “*to pin (an item)*”, e conforme referimos na seção 3.3.2, um dos problemas de tradução deste tipo de expressões é a restrição combinatória em relação à forma com que a área utiliza a expressão. De forma que não se pode substituir um dos elementos por sinônimos.

Por fim, um pouco abaixo dos erros de sobregeneralização, estão os erros de “derivação errada”, como mostra a tabela 37.

Tabela 37  
*Derivação errada*

N ° do exemplo	LP (inglês)	LC (PB)	Sugestão de tradução da CL (PB)
50	reset email	email redefinir	e-mail de redefinição de senha

O exemplo 50 mostra a criação de uma CL inexistente na LC.

Esta análise mostrou que os erros mais comuns de CLs são erros de substituição e que são influenciados por importação da LP, porém, apenas menos de 4% abaixo ficaram os erros

de variedade linguística, influência da LC, mas que, conforme referido, de certa forma, também podem ser considerados erros interlíngua.

A seção a seguir apresenta as sugestões para a empresa tendo em vista a implementação de melhorias.

## 5.4 Sugestões para a empresa

Em forma de resumo, apresentamos as contribuições que esperamos dar para a empresa através desta dissertação:

- Apresentamos as características das CLs, expressões que não podem, via de regra, ser traduzidas literalmente. Apontamos todas as informações relevantes e os impactos que podem causar para a tradução.
- Propomos a implementação de um tipo de erros para a anotação das CLs na taxonomia, como “*collocational errors*” em Costa *et al.* (2015). Na Unbabel ainda não existia um tipo de erro para estas expressões no momento em que este trabalho foi realizado.
- Exploramos vários tipos de erros que as CLs podem causar, tanto na tradução como na anotação das mesmas, o que permite uma visão geral do fenômeno e pode contribuir para o desenvolvimento de técnicas para a correção automática dos diferentes tipos de erros possíveis.
- Indicamos os domínios das CLEs, o que pode ajudar a resolver problemas de ambiguidade.
- Apontamos a origem dos erros de CLs, tanto na perspectiva “interlíngua” quanto na “intra-língua”.
- Sugerimos o aprimoramento das *Guidelines* através do estabelecimento de critérios de anotação relativamente às CLs (identificação e classificação), pois a anotação deve ser da expressão toda e conforme o tipo de erro apropriado, e com exemplos por língua de

trabalho, visto que as CLs são arbitrárias e variam de língua para língua, além de nem sempre um conceito ser expresso através de uma CL em todas as línguas (como a CLE da área das comunicações: “retorno de chamada” (PB), que em inglês utiliza apenas uma unidade lexical “*callback*”).

- Contribuímos para a resolução dos erros de TA e de pós-edição em relação às CLs, através do glossário criado. Este glossário é muito relevante, pois a análise foi feita com base em dados empíricos, os resultados reais da empresa, e todas as expressões incluídas no glossário estavam traduzidas incorretamente no *corpus*.
- Finalmente, contribuímos para a detecção automática das CLs das variedades linguísticas “PB e PE”, também através do glossário anexo, bem como para a detecção por parte dos editores e anotadores humanos, uma vez que estas expressões são idiossincráticas e, portanto, sempre desafiadoras para falantes não nativos.

Este capítulo teve o objetivo de descrever, detalhadamente, todas as observações e os resultados obtidos ao longo da pesquisa, através de exemplos, tabelas e gráficos; assim como apresentar as sugestões para a empresa. A análise foi feita no âmbito das CLs, a fim de contribuir para a melhoria do processo de anotação da empresa e para a diminuição dos erros de TA e pós-edição destas unidades. Adiante está a seção final deste trabalho com as conclusões.

## 6 Conclusões

Este trabalho tratou dos problemas de tradução das CLs, no ramo da TA e pós-edição, e analisou o processo de anotação das mesmas. A pesquisa foi com base em traduções de inglês para PB.

A motivação para focar o nosso estudo nas CLs foi porque estas expressões possuem propriedades que são muito desafiadoras para a TA, estão em todo tipo de comunicação e em todas as línguas.

O objetivo geral foi atingido. Os problemas de tradução de CLs no *corpus* analisado foram compreendidos e foram apresentadas as soluções encontradas através dos exemplos investigados e do glossário criado.

Os objetivos específicos também foram atendidos. Primeiramente, foi feita a descrição do estado da arte relativo às CLs, apresentando as características, os tipos, os problemas e as estratégias de tradução das mesmas. Em seguida, foi feita a verificação na plataforma de anotação nos tipos “*Lexical Selection*”, “*Overly Literal*” e “*Wrong Language Variety*”, a identificação dos domínios das CLEs e então: i) análise do processo de anotação relativamente às CLs. A partir desta análise, identificamos que não havia um tipo de erros específico para as CLs na taxonomia da empresa e propomos, então, a inclusão de um tipo de erros para a anotação destas expressões, assim como a implementação de critérios de anotação, e ii) análise dos erros de tradução de CLs. Nesta análise, apontamos a origem dos erros de CL e apresentamos uma possível explicação. Por último, foi conseguida a sistematização de erros frequentes de CLs entre PB e PE em tarefas de pós-edição. Ao caracterizar as CLs e identificar os problemas que suas propriedades causam para a tradução, verificamos que um dos impactos são as diferentes variedades linguísticas. Contribuímos, assim, para a detecção automática (e humana) das variedades linguísticas PB e PE através do glossário com as soluções encontradas para cada variedade.

Desta forma, os nossos resultados mostraram os problemas envolvidos na tradução das CLs e na anotação dos erros na Unbabel, principalmente, a ausência de um tipo de erros específicos para estas expressões na plataforma “*Annotate*”, e de critérios de anotação para as mesmas, já que os resultados também comprovaram que a maior parte dos erros de tradução do *corpus* analisado estavam relacionados com as CLs.

Para futuros trabalhos, sugerimos a continuação do glossário. O glossário foi criado com base nos erros de CLs do *corpus* analisado. Este trabalho pode ser continuado de acordo com outros erros que possam surgir futuramente nas traduções da empresa e pode ser expandido para outras línguas e também para outras áreas de especialidade.

No geral, a experiência deste estágio foi crucial para o meu desenvolvimento pessoal, acadêmico e profissional na área de tradução. A reflexão sobre as CLs e os problemas que elas causam para a tradução estará sempre presente na minha carreira como tradutora. Do mesmo modo, ter a oportunidade de trabalhar na área de TA e em um ambiente multicultural foi muito enriquecedor. Além destes ganhos, o processo de escrita desta dissertação me possibilitou aprimorar ainda mais as minhas competências de avaliação crítica de qualidade, organização, planejamento e gestão de tempo. No entanto, a demora na entrega dos dados por parte da empresa e, posteriormente, a necessidade de anonimizá-los devido à mudança na legislação de proteção dos dados dos clientes causaram atraso no processo de desenvolvimento desta dissertação. Apesar disso, no geral, a experiência foi muito positiva e, certamente, acrescentará muito nas minhas oportunidades futuras.

Esperamos, com este trabalho, contribuir não só para a Unbabel, mas também para a área de tradução, já que a análise do ponto de vista linguístico, e não computacional, contribui não só para a área de TA, mas para qualquer outra área da tradução, que é uma forma significativa de comunicação cultural e interpessoal em que as CLs devem ser amplamente exploradas.



## 7 Referências Bibliográficas

- Ágel, V. (2004). Phraseologismus als (valenz)syntaktischer Normalfall. In K. Steyer (Ed.), *Conexões de palavras - mais ou menos firmes* (pp. 65–86). Berlin/ Nova Iorque: de Gryuter.
- Aisenstadt, E. (1979). Collocability Restrictions in Dictionaries. *ITL - International Journal of Applied Linguistics*, 45–46(1), 71–74. <https://doi.org/10.1075/itl.45-46.10ais>
- Alonso Ramos, M., Wanner, L., Vincze, O., Casamayor del Bosque, G., Veiga, N. V., Suárez, E. M., & González, S. P. (2010). Towards a Motivated Annotation Schema of Collocation Errors in Learner Corpora. In N. et al. Calzolari (Ed.), *Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC'10)* (pp. 3209–3214). Paris: ELRA.
- Altenberg, B. (1998). On the Phraseology of Spoken English: The Evidence of Recurrent Word-Combinations. In A. P. Cowie (Ed.), *Phraseology: Theory, Analysis, and Applications* (pp. 101–122). New York: Clarendon Press. Oxford.
- Bally, C. (1961). *Traité de stylistique française* (Vol. 1). Paris: Klincksieck.
- Berman, A. (1985). Translations and the Trials of the Foreign. (Translated by Lawrence Venuti). In L. Venuti (Ed.), *The Translation Studies Reader* (pp. 284–297). New York / London: Routledge. Retrieved from [https://translationjournal.net/images/e-Books/PDF\\_Files/The Translation Studies Reader.pdf](https://translationjournal.net/images/e-Books/PDF_Files/The Translation Studies Reader.pdf)
- Bevilacqua, C. R. (2004/2005). Fraseologia: perspectiva da língua comum e da língua especializada. *Revista Língua e Literatura*, 6 e 7(10/11), 73–86. Retrieved from [https://www.researchgate.net/publication/237806911\\_Fraseologia\\_perspectiva\\_da\\_lingua\\_comum\\_e\\_da\\_lingua\\_especializada](https://www.researchgate.net/publication/237806911_Fraseologia_perspectiva_da_lingua_comum_e_da_lingua_especializada)
- Bizarro, M. S. F. (2014). *Diferenças tradutivas entre o português europeu e o português do Brasil: algumas observações num contexto específico de tradução técnica*. Faculdade de

Letras da Universidade de Coimbra, Coimbra.

Burger, Harald. (2003). *Phraseologie. Eine Einführung am Beispiel des Deutschen* (2nd ed.).

Berlin: Erich Schmidt Verlag.

Burger, Harald. (2004). Phraseologie - Kräuter und Rüben? Traditionen und Perspektiven der Forschung. In K. Steyer (Ed.), *Wortverbindungen - mehr oder weniger fest* (pp. 19–40).

Berlin, Boston: De Gruyter. <https://doi.org/10.1515/9783110622768-003>

Camacho, B. F. (2008). *Estudo comparativo de expressões idiomáticas do português do Brasil e de Portugal e do Francês da França e do Canadá*. Dissertação de mestrado. Universidade Estadual Paulista.

Castilho, A. T. (2010). *Nova Gramática Do Português Brasileiro*. São Paulo: Editora Contexto.

Castro, I. (1996). Para uma história do Português Clássico. In *Actas do Congresso Internacional sobre o Português* (Vol. II, pp. 135–150). Lisboa: Associação Portuguesa de Linguística.

Colaço, M. (2014). Combinatórias lexicais. In *Introdução às linguagens especializadas (Powerpoint slides)*. Mestrado em tradução e interpretação de conferências - FLUL/IPM.

Colson, J. P. (2003). Corpus linguistics and phraseological statistics: a few hypotheses and examples. In H Burger, A. Hächi Buhofer, & G. Gréciano (Eds.), *Flut von Texten - Vielfalt der Kulturen : Ascona 2001 zur Methodologie und Kulturspezifik der Phraseologie* (pp. 47–59). Baltmannsweiler: Schneider Verlag Hohengehren. Retrieved from <https://dial.uclouvain.be/pr/boreal/en/object/boreal%3A112823>

Corpas Pastor, G. (1998). Criterios generales de clasificación del universo fraseológico de las lenguas, com ejemplos tomados del Español y del Inglés. In M. A. Ezquerro & G. Copas Pastor (Eds.), *Diccionarios, Frases, Palabras* (pp. 157–187). Málaga: Servicio de Publicaciones de la Universidad.

Corpas Pastor, G. (2001). En torno al concepto de colocación. In *Euskera* (Vol. XLVI, pp. 89–108). Málaga: Universidad de Málaga.

Costa, Â. (2013). A diferença entre provérbio e expressão idiomática - Ciberdúvidas da Língua Portuguesa. Retrieved July 19, 2019, from <https://ciberduvidas.iscte-iul.pt/consultorio/perguntas/a-diferenca-entre-proverbio-e-expressao-idiomatica/32210>

Costa, Â., Ling, W., Luís, T., Correia, R., & Coheur, L. (2015). A linguistically motivated taxonomy for Machine Translation error analysis. *Machine Translation*, 29(2), 127–161. <https://doi.org/10.1007/s10590-015-9169-0>

Costa, Â. M. P. (2017). *As colocações em corpora de PL2: identificação, classificação e análise de erros*. Tese de Doutoramento. Universidade Nova de Lisboa.

Cowie, A. P. (1981). The Treatment of Collocations and Idioms in Learners' Dictionaries. *Applied Linguistics*, 2(3), 223–235. Retrieved from [internal-pdf:/Cowie 1981\(b\).pdf](http://internal-pdf:/Cowie%5C1981(b).pdf)

Cowie, A. P. (1998). Phraseological Dictionaries: Some East-West Comparisons. In A. P. Cowie (Ed.), *Phraseology: Theory, Analysis, and Applications* (pp. 209–228). New York: Clarendon Press. Oxford.

Cowie, A. P. (Ed.). (2001[1998]). *Phraseology: Theory, Analysis and Applications*. Oxford: Clarendon Press.

Darbelnet, J. (1970). Dictionnaires bilingues et lexicologie différentielle. *Langages*, 5(19), 92–102. <https://doi.org/10.3406/lgge.1970.2594>

Elloumi, Z., Besacier, L., & Kraif, O. (2015). Integrating Multi-Word Expressions in Statistical Machine Translation. In Gloria Corpas Pastor, J. Monti, V. Seretan, & R. Mitkov (Eds.), *Workshop Proceedings: Multi-word Units in Machine Translation and Translation Technologies* (pp. 83–86). Málaga, Spain. Retrieved from [http://www.europhras2015.eu/workshopproceedings\\_def/](http://www.europhras2015.eu/workshopproceedings_def/)

- Firth, J. R. (1968). Linguistic Analysis as a Study of Meaning. In F. R. Palmer (Ed.), *Selected Papers of J. R. Firth 1952-59* (pp. 12–26). London and Harlow: Longmans.
- Fontenelle, T. (1994). What on Earth are Collocations ? An assessment of the ways in which certain words co-occur and others do not. *English Today*, 10(4), 42–48. Retrieved from [http://www.ugr.es/~inped/exploringnewpaths/collocations/What\\_on\\_Earth\\_are\\_Collocations.pdf](http://www.ugr.es/~inped/exploringnewpaths/collocations/What_on_Earth_are_Collocations.pdf)
- Forcada, M. L. (2017). Making sense of neural machine translation. *Translation Spaces. A Multidisciplinary, Multimedia, and Multilingual Journal of Translation*, 6(2), 291–309. <https://doi.org/10.1075/ts.6.2.06for>
- Frias, M. J. M. (1991). Pedagogia inter-cultural e formação de professores de português, língua estrangeira. In *Actas Seminário Internacional Português como Língua Estrangeira* (pp. 443–450). Macau: Fundação Macau. Instituto Português do Oriente.
- Gonçalves, A., & Raposo, E. B. P. (2013). Verbos leves. In *Gramática do Português* (II, pp. 1214–1218). Coimbra: Fundação Calouste Gulbenkian.
- Grossmann, F., & Tutin, A. (2002). Collocations régulières et irrégulières: esquisse de typologie du phénomène collocatif. *Revue Française de Linguistique Appliquée*, VII, 7–25.
- Hashemi, H. B., & Hwa, R. (2014). A Comparison of MT Erros and ESL Erros. In *9th Language Resources and Evaluation Conference (LREC)*. Pittsburgh, PA: University of Pittsburgh. <https://doi.org/10.1002/ejic.201101347>
- Hausmann, F. J. (1984). Wortschatzlernen ist Kollokationslernen. Zum Lehren u. Lernen franzoesischer Wortverbindungen. *Praxis Des Neusprachlichen Unterrichts*, 31(4), 395–407.
- Hausmann, F. J. (1985). Kollokationen im deutschen Wörterbuch. Ein Beitrag zur Theorie des lexikographischen Beispiels. In H. Bergenholtz & J. Mugdan (Eds.), *Lexikographie*

- und Grammatik* (pp. 118–129). Tübingen: Max Niemeyer Verlag.
- Hotopf, W. H. N. (1983). Lexical Slips of the Pen and the Tongue: What they tell us about Language Production. In B. Butterworth (Ed.), *Language Production. Development, Writing and Other Language Processes* (11th ed., pp. 147–199). Londres/ Nueva York: Academic Press.
- Hutchins, J. W. (2015). Machine translation: History of Research and Applications. In *Routledge Encyclopedia of Translation Technology*. Routledge. Retrieved from <http://www.hutchinsweb.me.uk/Routledge-2014.pdf>
- Jorge, G. (1997a). Despedir-se à francesa / filer à l'anglaise: Reflexões em torno da tradutologia das construções fraseológicas na perspectiva interlínguas. In *Polifonia nº 1* (pp. 1–11). Lisboa: Colibri.
- Jorge, G. (Coord.). (1997b). Tradutor Dilacerado: reflexão de autores contemporâneos sobre tradução. In *Colecção voz de Babel* (pp. 16–63). Lisboa: Colibri.
- Jorge, G. (2001). Algumas reflexões em torno das expressões idiomáticas enquanto elementos que participam na construção de uma identidade cultural. In *Polifonia nº 4*. Lisboa: Colibri.
- Jorge, G. (2002). Da palavra às palavras - Alguns elementos para a tradução das expressões idiomáticas. In *Polifonia nº 5* (pp. 1–15). Lisboa: Colibri.
- Koike, K. (2000). *Colocaciones léxicas en el español actual: análisis formal y léxico-semántico*. Tesis Doctoral. Departamento de Lengua Española y Lingüística General. Facultad de Filología. UNED. Universidad Nacional de Educación a Distancia. Retrieved from <https://dialnet.unirioja.es/servlet/tesis?codigo=40367>
- L'Homme, M.-C. (2000). Understanding specialized lexical combinations. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication*, 6(1), 89–109. <https://doi.org/10.1075/term.6.1.06hom>

- Lommel, A., Burchardt, A., & Uszkoreit, H. (2015). Multidimensional Quality Metrics Definition. Retrieved May 27, 2019, from <http://www.qt21.eu/mqm-definition/definition-2015-12-30.html#mqm-core>
- Lorente, M. (2001). Terminología y fraseología especializada: del léxico a la sintaxis. In G. Guerrero (Ed.), *Terminología* (Vol. 5). Málaga: Grupo IULATERM. Retrieved from [http://www.upf.edu/pdi/iula/merce.lorente/docums/fras\\_c01.pdf](http://www.upf.edu/pdi/iula/merce.lorente/docums/fras_c01.pdf)
- Lorente, M., Bevilacqua, C. R., & Estopà, R. (2002). El análisis de la fraseología especializada mediante elementos de la lingüística actual. In *Terminologia, Desenvolvimento e Identidade Nacional - Actas do VI Simpósio Ibero Americano de Terminologia* (pp. 647–665). Lisboa: Edições Colibri / ILTEC.
- Martins, A. F. T., Almeida, M. B., & Smith, N. A. (2013). Turning on the Turbo: Fast Third-Order Non-Projective Turbo Parsers. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL)* (pp. 617–622). Sofia, Bulgária. Retrieved from <http://www.aclweb.org/anthology/P13-2109%5Cnhttp://www.cs.cmu.edu/~nasmith/papers/martins+almeida+smith.acl13.pdf>
- McKeown, K. R., & Radev, D. R. (2000). Collocations. In R. Dale, H. Moisl, & H. Somers (Eds.), *A Handbook of Natural Language Processing* (Marcel Dek, pp. 1–19). New York: Detarment of Computer Science. Columbia University.
- Mel'čuk, I. (1998). Collocations and Lexical Functions. In A. P. Cowie (Ed.), *Phraseology: Theory, Analysis, and Applications* (pp. 23–53). New York: Clarendon Press. Oxford.
- Misri, G. (1990). *Études traductologiques en hommage à Danica Seleskovitch*. Paris: Minard.
- Monti, J., & Todirascu, A. (2015). Multiword Units Translation Evaluation in Machine Translation: another pain in the neck? In G. Corpas Pastor, R. Mitkov, J. Monti, & V. Seretan (Eds.), *Workshop on Multi-word Units in Machine Translation and Translation Technology* (pp. 24–26). Málaga: LEXYTRAD, Research Group in Lexicography and

Translation Distribution.

- Ortiz Alvarez, M. L. (2000). *Expressões idiomáticas do português do Brasil e do espanhol de Cuba: Estudo contrastivo e implicações para o ensino de português como língua estrangeira*. Tese de Doutorado. Universidade Estadual de Campinas.
- Ortiz Alvarez, M. L. (2009). A motivação metafórica nas expressões idiomáticas como parte do patrimônio cultural dos povos latino-americanos. *Contextos, Estudios de Humanidades y Ciencias Sociales*, 21, 21–37.
- Ortiz Alvarez, M. L. (2012). Estudos fraseológicos no Brasil: Estado da arte. In *Tendências atuais na pesquisa descritiva e aplicada em fraseologia e paremiologia* (Volume I, pp. 355–375). Campinas, SP: Pontes Editores.
- Poulsen, S. (2005). *Collocations as a language resource: A functional and cognitive study in English phraseology*. Ph.d. dissertation. University of Southern Denmark.
- Sanromán, Á. I. (2000). *A unidade lexicográfica. Palavras, Colocações, Frasemas, Pragmatemas*. Dissertação de Doutoramento. Universidade do Minho, Braga. Retrieved from [https://repositorium.sdum.uminho.pt/bitstream/1822/4573/1/A\\_Unidade\\_Lexicografica.pdf](https://repositorium.sdum.uminho.pt/bitstream/1822/4573/1/A_Unidade_Lexicografica.pdf)
- Silva, R. V. M. (2013). O português do Brasil. In *Gramática do Português* (Vol. I, pp. 145–154). Coimbra: Fundação Calouste Gulbenkian.
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Testa, I. (2018). *Quality in Human Post-Editing of Machine-Translated Texts: Error Annotation and Linguistic Specifications for Tackling Register Errors*. Relatório de estágio de mestrado. Faculdade de Letras da Universidade de Lisboa, Lisboa.
- Unbabel. (2018a). *Unbabel*. (não disponível publicamente).
- Unbabel. (2018b). *Unbabel Annotation Guidelines*. Annotate. (não disponível publicamente)

(Vol. 1.2).

United Language Group. (2018). Statistical Vs. Neural Machine Translation. Retrieved May 23, 2019, from <https://unitedlanguagegroup.com/blog/statistical-vs-neural-machine-translation/>

Valente, R. (2005). Por que razão a combinatória lexical não pode ser um princípio para se considerar verbos e adjetivos como sendo especializados? In M. Doria & ILTEC (Eds.), *3ª Conferência Internacional de Terminologia Marítima* (pp. 176–186). Instituto de Linguística Teórica e Computacional.

Valente, R. S. (2000). Diferenças e similaridades colocacionais entre o português brasileiro e o português europeu - Estudo baseado na função lexical da Teoria Sentido-Texto. In *Cadernos do Congresso Nacional de Linguística e Filologia (CNLF)* (Vol. IV). Rio de Janeiro: Universidade do Rio de Janeiro.

Vilela, M. (2002). As expressões idiomáticas na língua e no discurso. In *Actas do Encontro Comemorativo dos 25 anos do Centro de Linguística da Universidade do Porto* (pp. 159–189). Porto: Centro de Linguística da Universidade do Porto. Faculdade de Letras da Universidade do Porto.

## Dicionários

“Arquivar”. *Michaelis*, <https://michaelis.uol.com.br/moderno-portugues/busca/portugues-brasileiro/arquivar/>. Consultado no dia 16/07/19.

“File”. *Cambridge Dictionary*, <https://dictionary.cambridge.org/pt/dicionario/ingles/file>. Consultado no dia 16/07/19.

“Head for”. *Collins Dictionary*, <https://www.collinsdictionary.com/dictionary/english/head-for>. Consultado no dia 30/05/19.



## Anexo - Glossário

Este glossário apresenta soluções de tradução para as CLs que estavam traduzidas incorretamente no *corpus* deste trabalho, abrangendo tanto expressões da língua comum quanto da linguagem especializada (nas áreas de TI, Finanças, Documentação, Comunicações, Comércio, Transportes, Entretenimento, além de saudações profissionais e CLs características dos e-mails).

Inclui entradas em que o conceito não é uma CL em uma língua, mas é na outra, tal como é o caso de “*screenshot*”, em inglês, que corresponde a “captura de tela” em PB e “captura de ecrã” em PE.

Para a tradução das CLs, principalmente a TA, o glossário é uma ferramenta muito útil, visto que estas expressões nem sempre podem ser traduzidas literalmente. Assim, pretendemos, com este glossário, contribuir para a resolução dos erros de CLs e para a detecção das variedades linguísticas “PB e PE”, em relação a estas expressões.

As entradas estão em ordem alfabética. Nas duas primeiras colunas, estão as CLs da LP e das LC (no caso do PE, só foram adicionadas as traduções que são diferentes da expressão no PB). Nas três colunas seguintes, são apresentados, respectivamente, a definição da CL, o domínio e as fontes (seja da consulta da CL na LP ou do equivalente utilizado na LC) e por fim, na última coluna estão colocadas algumas observações. Dado que a criação de um glossário é um processo em constante atualização e o tempo de execução desta dissertação era limitado, existem alguns espaços por preencher que, assim como a inclusão de novas CLs, podem ser constantemente implementados.

As fontes utilizadas foram: os dicionários monolíngues *Collins Dictionary*, *Macmillan Dictionary*, *Urban Dictionary*, *The free dictionary of Idioms*, Priberam, o dicionário online de português “Dicio”; o vocabulário ortográfico da língua portuguesa (VOLP) da academia brasileira de

letras, as *Unbabel Language Guidelines – Portuguese (BR)* e os *corpora* paralelos “*Linguee*”. Para as CLEs, além destas fontes em alguns casos, foram utilizadas algumas bases de dados especializadas, como IATE (*InterActive Terminology for Europe*), *Microsoft Terminology Collection*, o glossário ISACA e sites *online*.

Para além destas fontes, houve a preocupação de procurar, tanto para o PB como para o PE, atestações das CLs em textos dos respectivos domínios.

<b>Língua de partida (inglês)</b>	<b>Língua de chegada (PB)</b>	<b>Língua de chegada (PE)</b>	<b>Definição</b>	<b>Domínio</b>	<b>Fonte</b>	<b>Observações</b>
Acceptable Use Policy	Política de Uso Aceitável	Política de Utilização Aceitável / Política de Uso Aceitável	Uma declaração emitida por um serviço on-line que indica em quais atividades os usuários podem ou não participar enquanto estão conectados ao serviço.	TI	Microsoft Terminology Collection, IATE	Obs: No IATE também “Política de Uso Aceitável” para o PE
access your account	acessar sua conta	aceder à sua conta		TI	Microsoft Terminology Collection	
account details	dados da conta			TI	Linguee	
Account Settings	Configurações de Conta	Definições de Conta		TI	Microsoft Terminology Collection	
account team	equipe de contas	equipa de contabilidade		Finanças	Linguee	
Apologies for any inconveniences caused.	Desculpa por qualquer inconveniente / transtorno causado.	Pedimos desculpa por qualquer inconveniente / transtorno causado.		Saudações profissionais		
as soon as possible / as fast as we can	o mais rápido possível	o mais rapidamente possível		língua comum		
at the moment	no momento	de momento		língua comum	Ciberdúvidas	

ATM	caixa eletrônico / caixa automático	caixa multibanco / multibanco	É um dispositivo eletrônico que permite que clientes de um banco saquem dinheiro e verifiquem o saldo de suas contas bancárias sem precisar de um funcionário do banco. São os principais equipamentos de automação bancária. Muitos terminais também permitem depósito de dinheiro ou cheques, consulta de saldos, extratos, transferências, pagamentos, agendamentos, empréstimos, investimentos, financiamentos e desbloqueio, etc.	Finanças	IATE, Dicionário informal	
Bank Statement	extrato bancário	extrato de conta / extrato bancário	Relatório contendo informações sobre a movimentação e o saldo de uma conta bancária.	Finanças	Microsoft Terminology Collection, IATE	
based on chance	de modo aleatório / por acaso			Entretenimento - jogos	Linguee	
Best Regards / Regards / Sincerely / Cheers / All the best	Atenciosamente, Cordialmente	Cordialmente, Atentamente, Com os (meus / nossos) melhores cumprimentos		Saudações profissionais	<a href="https://catracalivre.com.br/carreira/8-dicas-para-escrever-bem-seus-e-mails-corporativos/">https://catracalivre.com.br/carreira/8-dicas-para-escrever-bem-seus-e-mails-corporativos/</a> , <a href="https://ciberduvidas.iscte-iul.pt/outros/diversidades/5-dicas-para-escrever-um-e-">https://ciberduvidas.iscte-iul.pt/outros/diversidades/5-dicas-para-escrever-um-e-</a>	

					mail-profissional-eficaz/3589	
blog post	postagem de blog	mensagem de blogue / publicação de blogue	Uma entrada individual em um blog.	TI	Microsoft Terminology Collection	
callback	retorno de chamada	chamada de retorno	Nas telecomunicações, um retorno de chamada ocorre quando o originador de uma chamada recebe imediatamente outra chamada como resposta.	Comunicações	Microsoft Terminology Collection, IATE	
card delivery address confirming document	documento de confirmação do endereço de entrega do cartão			Documentação		
cards reissue	reemissão de cartões			Finanças		
Check both the (PRODUCT) and phone ports for debris	verifique se há resíduos nas portas do telefone / entradas do telefone			Comunicações		
clear off	se resolver			língua comum		

click here	clique aqui			TI	Priberam e Dicio	
click through" on the image	clicar na imagem			TI	Linguee	
close the application	fechar o aplicativo	fechar a aplicação		TI	Microsoft Terminology Collection	
close this ticket	fechar este tíquete (de suporte) / este ticket	fechar este pedido (de suporte) / este <i>ticket</i>		TI	Microsoft Terminology Collection	
come back to us	fale conosco novamente	fale connosco novamente		Comunicações		
complete the return	concluir a devolução			Comércio	Microsoft Terminology Collection	
contactless debit card	cartão de débito por aproximação	cartão de débito sem contacto	Um cartão com chip com interface de frequência de rádio sem contato, que tem a funcionalidade (mas às vezes não exatamente a arquitetura) de um cartão inteligente, ou seja: memória multi-aplicativos, procedimentos de segurança como autenticação e criptografia e recursos de cálculo para valores armazenados no cartão.	TI e processamento de dados	<a href="https://www.mastercard.com.br/pt-br/consumidores/beneficios/contactless.html">https://www.mastercard.com.br/pt-br/consumidores/beneficios/contactless.html</a> , IATE	
contact details	contatos / informações de contato	contactos		Comunicações	IATE	

could you write in an email from the email you'd like to verify						
credit card pre-authorization check	verificação de autorização de cartão de crédito			Finanças	Linguee	
debug logs	logs de depuração	registos de depuração		TI	Microsoft Terminology Collection	
delete cookie	excluir cookie	eliminar cookie		TI	Microsoft Terminology Collection	
display results	exibir resultados	apresentar resultados		TI	Microsoft Terminology Collection	
Do <b>let us know</b> if you have further question.	Em caso de dúvidas adicionais, por favor, avise-nos / nos avise.	Em caso de dúvidas adicionais, por favor, avise-nos.		Saudações profissionais	Linguee	“Avise-nos” é uma opção mais formal no PB.
Do not hesitate to contact us for any other questions or doubts.	Não hesite em entrar em contato em caso de dúvidas ou quaisquer esclarecimentos adicionais.	Não hesite em entrar em contacto em caso de dúvidas ou quaisquer esclarecimentos adicionais.		Saudações profissionais		
Download your design	fazer o download do / baixar o design	Descarregar / transferir		TI	Microsoft Terminology Collection	

driving license / driver's license	Carteira Nacional de Habilitação (CNH), Carteira / carta de motorista, Carteira / carta de habilitação	carta de condução	Documento que atesta a aptidão de um cidadão para conduzir veículos automotores terrestres.	Documentação	Linguee	
drop your package off	entregar o pacote			Comércio, Transporte	Linguee	
electoral roll card	título eleitoral	cartão de eleitor		Documentação		
electricity bill	conta de energia elétrica / de luz	fatura de eletricidade	Uma conta para o dinheiro devido pela eletricidade usada.	Finanças	<u>ENEL / EDP,</u> <a href="https://www.thefreedictionary.com/electric+bill">https://www.thefreedictionary.com/electric+bill</a>	
email address	endereço de e-mail	endereço de e- mail / endereço de correio eletrónico / endereço eletrónico	Endereço graças ao qual um cibernauta pode comunicar por e-mail com outros cibernautas (receber e enviar mensagens).	TI, Comunicações	Microsoft Terminology Collection, VOLP, IATE	
entry permit sticker	carimbo de autorização de entrada			Documentação		
escalate this issue (problem escalation procedure)	encaminhar este problema (procedimento de encaminhamento de problemas)		O processo de encaminhar um problema do pessoal júnior para sênior e, finalmente, para níveis mais altos de gerenciamento. Observação: O procedimento de encaminhamento de	TI	ISACA	Obs: em “Microsoft Terminology Collection” aparece a opção “escalar o problema”.



			problemas é frequentemente usado no gerenciamento de help desk, quando um problema não resolvido é encaminhado até a cadeia de comando, até ser resolvido.			
estimated arrival	previsão de chegada	hora prevista de chegada, hora de chegada prevista, tempo estimado de chegada, hora estimada de chegada	Hora que o responsável pelo meio de transporte estima e anuncia como sendo a de chegada a qualquer local, incluindo uma estação de pilotos ou um cais de atracação.	Transporte	Linguee / IATE	
European Economical Area	Área Econômica Europeia (EEA)	Espaço Económico Europeu (EEE) / Área Económica Europeia (AEE)	Uma área geográfica criada por instâncias europeias, para permitir a livre circulação dos bens, dos serviços, das pessoas e dos capitais; as quatro liberdades fundamentais, livre circulação essa, efetuada dentro do mercado interno da União Europeia.	Finanças	IATE, Wikipedia, Linguee, <a href="https://veja.abril.com.br/noticias-sobre/area-economica-europeia/">https://veja.abril.com.br/noticias-sobre/area-economica-europeia/</a>	
failed payments	pagamentos não concretizados		Uma transação de pagamento que foi recusada pela emissora do cartão de crédito ou banco do seu cliente.	Finanças	Linguee, <a href="https://www.ezypay.com/features/failed-payment-handling">https://www.ezypay.com/features/failed-payment-handling</a>	

file a chargeback request	apresentar / abrir um pedido de estorno			Finanças	IATE, Linguee	Não confundir com “arquivar”, que não é um equivalente na área financeira na LC.
flight cases	flight cases		Contentor de transporte construído especificamente para proteger instrumentos musicais, equipamento cinematográfico, equipamento de produção de áudio e iluminação, propriedades ou outro equipamento sensível quando este tiver de ser transportado, ou frequentemente atirado pelo pessoal de manutenção de bagagens do aeroporto.		<a href="https://www.universalcases.com.br/">https://www.universalcases.com.br/</a>	
flight data	dados de voo	dados de voo / elementos de voo	Dados relativos à circulação efetiva ou prevista de aeronaves, normalmente apresentados sob uma forma codificada ou abreviada.	TI e processamento de dados / Transporte	Linguee, IATE	
follow up	Fazer um follow up, fazer o acompanhamento, acompanhar, seguir	dar seguimento / seguir	Fazer investigações adicionais sobre um contato, com base nas informações recebidas.	TI	Microsoft Terminology Collection	
free shipment	envio gratuito / frete gratuito			Transporte	Linguee	

full credit card information	dados completos do cartão			TI		
FX fee / foreign transaction fee						
get caught in the mix	ser encontrado		Incluso.	língua comum	Urban dictionary	
gift card	cartão-presente	cartão oferta / cartão de oferta / cartão de presente /vale de oferta	Um cartão físico ou eletrônico com um valor armazenado que os clientes podem usar para comprar aplicativos e itens semelhantes.	Comércio	Microsoft Terminology Collection, Linguee	Menos frequente: vale de compras
give us a call	entre em contato conosco	entre em contacto conosco		Comunicações	Linguee	
go a long way	contribui fortemente para		É um fator importante para alcançar algo desejado.	língua comum	Collins, Linguee	
go to 'Login'	clique em "login" / acesse	carregue em "entrar" / "iniciar sessão"		TI		
Government Letters	carta de serviços			Documentação	<a href="http://fazenda.gov.br/carta-de-servicos">http://fazenda.gov.br/carta-de-servicos</a>	
have (your ID document) ready	tenha (o seu documento de identificação) em mãos			língua comum	Linguee	
have you had any discussions	já entrou em contato com	já entrou em contacto com		língua comum		

Hello / Hi / Hey	Olá / Prezado(a) / Caro(a) Senhor(a)	Olá / Prezado(a) / Estimado(a) / Caro(a)	Formas de cumprimento.	Saudações profissionais	<a href="https://catracalivre.com.br/carreira/8-dicas-para-escrever-bem-seus-emails-corporativos/">https://catracalivre.com.br/carreira/8-dicas-para-escrever-bem-seus-emails-corporativos/</a> , <a href="https://ciberduvidas.iscte-iul.pt/outros/diversidades/5-dicas-para-escrever-um-email-profissional-eficaz/3589">https://ciberduvidas.iscte-iul.pt/outros/diversidades/5-dicas-para-escrever-um-email-profissional-eficaz/3589</a>	
help you accordingly	ajudá-lo neste sentido			língua comum		
I am glad you are getting back to us	Obrigado(a) pelo retorno / pela resposta			Saudações profissionais	Linguee	
I am looking forward to hearing from you soon / I look forward to hearing back from you	Fico no aguardo / aguardo retorno	fico a aguardar / aguardo resposta		Saudações profissionais	<a href="https://ciberduvidas.iscte-iul.pt/consultorio/perguntas/uso-de-aguardo-no-portugues-europeu/29024">https://ciberduvidas.iscte-iul.pt/consultorio/perguntas/uso-de-aguardo-no-portugues-europeu/29024</a>	
I am unsure	não tenho certeza	não tenho a certeza		língua comum	Linguee	
I do understand your disappointment this situation may have caused.	compeendo o inconveniente que essa situação possa ter causado.			Saudações profissionais		
I have done a check with	Verifiquei / chequei com			língua comum		
I have raised this with them again (raise something with someone)	Levantei esta questão com eles novamente		Levantar uma questão com alguém.	língua comum	<a href="https://idioms.thefreedictionary.com/raise+with">https://idioms.thefreedictionary.com/raise+with</a>	

I understand	compreendo / entendo que	percebo / compreendo / entendo que		língua comum	Linguae	
I understand this can be really frustrating				língua comum		
<b>I would like to encourage you</b> to click on the link	por favor, clique no link			língua comum		
I've raised your comments (to the)	Encaminhei os seus comentários (para)			língua comum		
I'd be happy to help.	Será um prazer ajudá-lo.	Terei muito gosto em ajudá- lo.		Saudações profissionais	Linguae	
If there is anything else we can help you with, or if you have any further queries, <b>please do not hesitate to contact us.</b>	Por favor, não hesite em nos contatar / entrar em contato em caso de dúvidas ou quaisquer esclarecimentos adicionais.	contactar / entrar em contacto		Saudações profissionais		
<b>If there is nothing else</b> let me know and I'll mark this ticket as solved.	Se estiver resolvido, por favor, avise-me / me avise...	Se estiver resolvido, por favor, avise- me...				
<b>I'll go ahead</b> (and close this ticket for you).				língua comum		

In case of any further questions do not hesitate to contact us back. / In case of any other questions feel free to contact us back.	Em caso de dúvidas adicionais, não hesite em contatar-nos / nos contatar / entrar em contato.	Em caso de dúvidas adicionais, não hesite em contactar-nos / entrar em contacto.		Saudações profissionais		
keep in mind	lembre-se de que / observe que	tenha em atenção	Pensar ou lembrar alguém ou algo.	língua comum	Linguee, Microsoft Terminology Collection, <a href="https://idioms.thefreedictionary.com/keep+in+mind">https://idioms.thefreedictionary.com/keep+in+mind</a>	
Kindly note that	Lembre-se de que / observe que	tenha em atenção		língua comum	Linguee, Microsoft Terminology Collection	
<b>Let me know</b> if that resolves your issue,	Por favor, avise-me / me avise se isso resolve o problema.	Por favor, avise-me se isso resolve o problema.		Saudações profissionais	Linguee, Unbabel Language Guidelines – Portuguese (BR)	“Avise-me” é uma opção mais formal no PB.
Let me know if you have further questions.	Em caso de dúvidas adicionais, por favor, entre em contato.	Em caso de dúvidas adicionais, por favor, entre em contacto.		Saudações profissionais	Microsoft Terminology Collection	
Let me know if you still have difficulties.	Por favor, avise-nos / nos avise se o problema persistir.	Por favor, avise-nos se o problema persistir.		Saudações profissionais	Unbabel Language Guidelines – Portuguese (BR)	“Avise-nos” é uma opção mais formal no PB.
linking code	código de vinculação	código de associação		TI	Microsoft Terminology Collection	

log files	arquivos de log / log de dados	ficheiros de relatório / ficheiro de registo / diário / ficheiro cronológico / ficheiros de log	Arquivo que mantém um registro cronológico dos eventos que ocorrem num aplicativo informático ou num sistema informático e permite analisar a respectiva atividade.	TI	Microsoft Terminology Collection, IATE, Wikipedia	
log in / sign in	entrar / fazer login	iniciar sessão	Processo de acessar um sistema informático, habitualmente pela introdução da identificação do usuário e de uma senha.	TI	Microsoft Terminology Collection, IATE	
Look forward.	Fico no aguardo / aguardo retorno.	Fico a aguardar. / Aguardo resposta.		Saudações profissionais	<a href="https://ciberduvidas.iscte-iul.pt/consultorio/perguntas/uso-de-aguardo-no-portugues-europeu/29024">https://ciberduvidas.iscte-iul.pt/consultorio/perguntas/uso-de-aguardo-no-portugues-europeu/29024</a>	
Looking forward to your reply,	Fico no aguardo / aguardo retorno.	Fico a aguardar. / Aguardo resposta.		Saudações profissionais	<a href="https://ciberduvidas.iscte-iul.pt/consultorio/perguntas/uso-de-aguardo-no-portugues-europeu/29024">https://ciberduvidas.iscte-iul.pt/consultorio/perguntas/uso-de-aguardo-no-portugues-europeu/29024</a>	
make a video selfie				TI		
<b>make a withdrawal</b> to your account	<b>sacar / retirar dinheiro</b> da conta / <b>fazer um saque</b>	<b>levantar dinheiro</b> da conta / <b>fazer um levantamento</b>		Finanças	<a href="https://www.bb.com.br/pbb/pagina-inicial/voce/produtos-e-servicos/contas/todos-os-servicos/saque-movel">https://www.bb.com.br/pbb/pagina-inicial/voce/produtos-e-servicos/contas/todos-os-servicos/saque-movel</a> , IATE	
make an edit	faça uma edição			TI		

May I ask how exactly can I be of help?				língua comum		
mobile app	aplicativo móvel	aplicação móvel	Um software desenvolvido para ser instalado em smartphones, telefones celulares e outros dispositivos eletrônicos móveis.	TI	Microsoft Terminology Collection	
Mobile Banking service	serviço de internet banking no celular / serviço de banco pelo celular / banco móvel / mobile banking	serviço bancário móvel / banca móvel	São ferramentas que disponibilizam alguns serviços tipicamente bancários através de dispositivos móveis, como um celular.	Finanças, TI	Linguee, IATE, Wikipedia	
mobile operator	prestadora / operadora de telefonia móvel	operador de rede móvel	Operadora de infraestrutura de rede móvel que serve de suporte à transmissão e fornecimento de serviços de comunicação por rádio. As atividades das operadoras de redes móveis, na maioria dos casos, integram igualmente funções de provedor de serviços móveis (prestação de serviços diretos aos usuários finais) no âmbito do conjunto de suas atividades.	Comunicações	<a href="http://www.anatel.gov.br/dados/controle-de-qualidade/controle-telefonia-movel">http://www.anatel.gov.br/dados/controle-de-qualidade/controle-telefonia-movel</a> , IATE	



national ID, ID card, identity card	Documento de identificação nacional / cédula de identidade, carteira de identidade (RG)	bilhete de identidade (BI) / cartão de cidadão	Documento bastante para provar a identidade civil do seu titular perante quaisquer autoridades, entidades públicas ou privadas.	Documentação, Direito	<a href="http://www.sejusp.ms.gov.br/carteira-de-identidade/IATE">http://www.sejusp.ms.gov.br/carteira-de-identidade/IATE</a>	
no older than 6 months	com menos de 6 meses / não deve ter mais de seis meses			Documentação		
nominated account	conta vinculada / adicionada / cadastrada	conta designada	Uma conta de algum banco que o titular utiliza para fazer e receber pagamentos eletrônicos.	Finanças	Linguee, IATE, <a href="https://www.paypal.com/br/smarthelp/article/como-adiciono-uma-conta-banc%C3%A1ria-%C3%A0-minha-conta-do-paypal-faq686">https://www.paypal.com/br/smarthelp/article/como-adiciono-uma-conta-banc%C3%A1ria-%C3%A0-minha-conta-do-paypal-faq686</a>	
on hold (a case)	em espera	em espera / em suspenso	O estado de um processo operacional que foi suspenso ou o status de um documento e produto que participam de um processo que está suspenso.	TI	Microsoft Terminology Collection	
on your behalf	em seu nome		Para o benefício, interesse ou apoio de alguém; como um agente, representante ou no lugar de alguém.	língua comum	<a href="https://idioms.thefreedictionary.com/on+your+behalf">https://idioms.thefreedictionary.com/on+your+behalf</a>	
Online Store	loja online / loja virtual		Um site ou aplicativo por meio do qual os bens ou serviços são vendidos pela Internet.	TI	Microsoft Terminology Collection	

open the app / application	abra o aplicativo	abra a aplicação		TI	Microsoft Termonology Collection	
opening hours	horário de funcionamento / horário de atendimento	horário de funcionamento / horário de atendimento / horários de abertura		Finanças	IATE	
optional details	dados opcionais			TI		
order	fazer o pedido		Solicitar a entrega de, pedir para trazer.	Comércio	<a href="http://www.logosdictionary.org/index.php">http://www.logosdictionary.org/index.php</a>	
original image source	imagem de origem					
Our <b>support line</b> number is <b>operating</b> Monday through Friday	A nossa linha de apoio funciona de segunda à sexta-feira.			Comunicações		
passionate volunteer				língua comum		
payment modes	formas de pagamento / métodos de pagamento			Finanças		
payment via credit card	pagamentos com cartão de crédito			Finanças	Linguee	

photo ID / photo identification	documento de identificação com foto		É um documento de identidade que inclui uma fotografia do titular, normalmente apenas o rosto.	Documentação	Wikipedia	
pick up the shipment	retirar o frete	levantar o frete		Transporte	Linguee	
pin	marcação	marcador	Um item, como um filme, um jogo ou um aplicativo que está fixado em uma determinada área da interface do usuário, de forma que está sempre acessível naquela área.	TI	Microsoft Terminology Collection	
pin (an item)	fixar (um item)	afixar (um item)	Fixar um item, como um bloco, biblioteca, filme, jogo ou aplicativo, em uma determinada área da interface do usuário, para que esteja sempre acessível nessa área (por exemplo, tornar uma biblioteca sempre visível no painel de navegação).	TI	Microsoft Terminology Collection	
Pin removal	remoção de marcações	remoção de marcadores		TI		
Please be aware	Esteja ciente de que / Lembre-se que / É importante lembrar que	Tome cuidado / Note que / Tenha em atenção que / Certifique-se de que			Microsoft Terminology Collection	

Please clarify	Por favor, esclareça			língua comum	Linguee	
please do write back in case the problem persists	por favor, entre em contato novamente se o problema persistir	por favor, entre em contacto novamente se o problema persistir		Saudações profissionais		
please feel free to get back to us	por favor, não hesite em entrar em contato	Por favor, não hesite em entrar em contacto				
please feel free to get in touch	por favor, não hesite em entrar em contato	Por favor, não hesite em entrar em contacto		Saudações profissionais	Linguee	
Please let me know	Por favor, me avise / avise-me	Por favor, avise-me		Saudações profissionais	Linguee, Unbabel Language Guidelines – Portuguese (BR)	
Please let me know how that goes for you.	Por favor, nos avise como foi / se deu certo	Por favor, depois diga-me como correu / se resultou		língua comum		
please let me know if there is anything else we can do to assist!				Saudações profissionais		
Please let me know if you have any other <b>questions or concerns</b> .	Em caso de dúvidas ou esclarecimentos adicionais, por favor, me avise / avise-me / entre em contato.	Em caso de dúvidas ou esclarecimentos adicionais, por favor, avise-me / entre em contacto.		Saudações profissionais	Microsoft Terminology Collection	

Please let me know if you have any troubles with this and <b>I will be more than happy to help.</b>	Será um prazer ajudá-lo.	Terei muito gosto em ajudá-lo.		Saudações profissionais	Linguee, <a href="https://ciberduvidas.iscte-iul.pt/consultorio/perguntas/ter-muito-gosto-em-vs-ter-o-gosto-de/34816">https://ciberduvidas.iscte-iul.pt/consultorio/perguntas/ter-muito-gosto-em-vs-ter-o-gosto-de/34816</a>	
Please let us know any of the following details	Por favor, nos informe algum dos seguintes dados	Por favor, informe-nos algum dos seguintes dados		língua comum		
Please note that	observe que			língua comum		
Please pay attention that	Por favor, certifique-se de que / preste atenção que			língua comum	Linguee	
post messages	postar mensagens	publicar mensagens		TI	Microsoft Terminology Collection	
proof of address	comprovante de residência / endereço	comprovativo de morada	Qualquer documento que pode ser utilizado para comprovar que o indivíduo possui uma residência fixa (seja ela própria ou alugada).	Documentação	Linguee	
prospect customers	cliente em potencial			Finanças		
Publishing Administration Dashboard	Painel de Gerenciamento de Publicações			TI		

query log file	arquivo de log de consultas	ficheiro de registo de consultas		TI	Microsoft Terminology Collection	
reach out to us	contatar-nos / nos contatar / entrar em contato	contactar-nos / entrar em contacto		Saudações profissionais		
receive delivery (of your recent (COMPANY) order).				Comércio		
receive your return	receber a devolução			Comércio	Linguee	
reset email	email de redefinição de senha			TI	Microsoft Terminology Collection	
rest assured	Pode ter certeza / garantimos que	Pode a ter certeza / garantimos que	Ter certeza de que algo vai acontecer.	língua comum	Cambridge Dictionary, Linguee	
return department	departamento de devoluções			Comércio, Transporte	Linguee	
return items	devolver itens			Comércio	Linguee	
return label	etiqueta de devolução	etiqueta de retorno	Uma etiqueta postal pré-paga e pré-endereçada que permite aos clientes devolver os produtos adquiridos ao vendedor.	Comércio	Microsoft Terminology Collection, IATE, <a href="https://www.stamps.com/ups/return-shipping-label/">https://www.stamps.com/ups/return-shipping-label/</a>	
return your order	devolver o pedido			Comércio		
review the application	analisar o aplicativo	analisar a aplicação	Fazer o resumo dos aplicativos de sensoriamento remoto.	Comunicações	IATE	

run the search	Executar a pesquisa	Fazer a pesquisa / executar a pesquisa		TI		
save the file	salvar / gravar o arquivo	guardar / gravar o ficheiro	Armazenar (arquivos) em um computador ou em um dispositivo de armazenamento (como um CD ou unidade flash).		<a href="https://www.merriam-webster.com/dictionary/save">https://www.merriam-webster.com/dictionary/save</a> , Microsoft Terminology Collection	
screenshot	captura de tela / screenshot	captura de ecrã	Uma imagem que reproduz toda ou parte da tela de um computador ou outro monitor.	TI	Microsoft Terminology Collection / Wikipedia	
scroll down	rolar para baixo	deslocar (ecrã) para baixo		TI	Microsoft Terminology Collection	
search engine	mecanismo de pesquisa	motor de busca, motor de pesquisa, buscador	Programa que indexa o conteúdo de diferentes recursos da Internet, designadamente de arquivos e sites da Web, e que permite ao internauta pesquisar informação segundo diferentes parâmetros, servindo-se de senhas, e acessar à mesma.	TI	Microsoft Terminology Collection, IATE	
secondary ID	documento secundário			TI, processamento de dados	Microsoft Terminology Collection, IATE	
sensitive information	informações confidenciais / sigilosas	informação sensível, informações sensíveis / confidenciais / sigilosas	Informação cuja divulgação, alteração, destruição ou perda é suscetível de prejudicar pessoas ou bens.	Finanças	Microsoft Terminology Collection, IATE	

Settings menu	menu Configurações	menu Definições		TI	Microsoft Terminology Collection	
should not be older than 3 months	não deve ter mais de 3 meses			Documentação		
Should you have any questions, <b>will be glad to answer.</b>	será um prazer respondê-lo	terei muito gosto em responder		Saudações profissionais	Linguee, <a href="https://ciberduvidas.iscte-iul.pt/consultorio/perguntas/ter-muito-gosto-em-vs-ter-o-gosto-de/34816">https://ciberduvidas.iscte-iul.pt/consultorio/perguntas/ter-muito-gosto-em-vs-ter-o-gosto-de/34816</a>	
sign up	Inscriver-se / criar conta / assinar / cadastrar-se / registro / inscrição	inscrever-se / aderir / registrar / inscrição / efectuar o registro / iniciar sessão	Se inscrever em um serviço, geralmente resultando na criação de uma nova conta.	TI	Microsoft Terminology Collection	
sign up process	processo de inscrição			TI	Microsoft Terminology Collection	
<b>sorry to hear</b> you've had some trouble with...	sinto muito pelo problema com...	Lamento muito...		Saudações profissionais	Linguee, Unbabel Language Guidelines – Portuguese (BR)	
<b>Sorry to know</b> that you're encountering an issue...	Sinto muito pelo problema com...	Lamento muito...		Saudações profissionais	Linguee, Unbabel Language Guidelines – Portuguese (BR)	
Sorry you're having some trouble	Sinto muito que esteja tendo problemas...	Lamento muito...		Saudações profissionais		
Spam folder / Junk Mail folder	pasta Lixo Eletrônico	pasta Correio Publicitário Não Solicitado / pasta correio	Página onde os usuários podem armazenar anúncios não desejados e mensagens de e-mail.	TI	Microsoft Terminology Collection	



		eletrónico de lixo				
spammy pins	marcações de spam	marcadores de spam		TI	Microsoft Terminology Collection	
speed up the call	facilitar a ligação			Comunicações		
stay tuned	fique ligado / atento		Continuar prestando atenção para novas informações.	língua comum	Linguee, <a href="https://idioms.thefreedictionary.com/stay+tuned">https://idioms.thefreedictionary.com/stay+tuned</a>	
stuck in sign up process somewhere				TI		
submit a ticket	enviar um tíquete (de suporte)	submeter um pedido (de suporte)		TI	Microsoft Terminology Collection	
subtitle content	legendar conteúdo					
support line	linha de apoio / suporte	linha de apoio / atendimento		TI, Comunicações	Microsoft Terminology Collection, Linguee	
support number				TI		

support standing orders	atender ordens permanentes	atender ordens permanentes / ordens de domiciliação	Uma instrução que o titular de uma conta bancária ("o pagador") dá ao seu banco para pagar uma quantia fixa em intervalos de tempo regulares à conta de outro ("o recebedor").	Finanças	IATE	
Support team	equipe de suporte / apoio	equipa de suporte / apoio	Um grupo ou organização que é responsável por fornecer suporte técnico.	TI / Comunicações	Microsoft Terminology Collection, IATE	
switch over (the phone)	Transferir		Transferir eletronicamente um sinal de alguém ou alguma coisa para alguém ou outra coisa.	Comunicações	<a href="https://idioms.thefreedictionary.com/switch+over+to">https://idioms.thefreedictionary.com/switch+over+to</a>	
take a clear image	tirar uma foto nítida / imagem nítida	capturar uma imagem nítida			Linguee	
take a look	ver / dar uma olhada	dar uma olhadela / dar uma vista de olhos	Olhar de relance, olhar um pouco, olhar rapidamente.	língua comum	<a href="https://idioms.thefreedictionary.com/take+a+look+at">https://idioms.thefreedictionary.com/take+a+look+at</a> , <a href="#">Dicionário Informal</a>	
tap Share	tocar em compartilhar	tocar / carregar em partilhar		TI	Microsoft Terminology Collection	
thank you for contacting us!	Obrigado(a) por contactar-nos / nos contactar / entrar em contato	nos contactares / contactar		Saudações profissionais		

Thank you for <b>taking the time</b> to create an account with us today!	Obrigado(a) por criar uma conta conosco!	Obrigado(a) por criar uma conta conosco!		Saudações profissionais		
Thank you for <b>taking the time</b> to write in.	Obrigado(a) por entrar em contato / escrever / Obrigado(a) pelo e-mail	Obrigado(a) por entrar em contato / escrever / Obrigado(a) pelo e-mail		Saudações profissionais - formas de agradecimento		
Thank you for the update and <b>I'm really glad to hear things are moving along!</b>	Fico feliz por saber que as coisas estão indo bem.	Fico feliz por saber que as coisas estão a correr bem.		língua comum	Linguee	
Thank you for writing in to us	Obrigado(a) por entrar em contato	Obrigado(a) por entrar em contato		Saudações profissionais		
thank you for your understanding,	obrigado(a) pela compreensão			Saudações profissionais	Linguee	
Thanks for <b>connecting</b> with (COMPANY) Support	obrigado(a) por entrar em contato / contatar com o suporte	obrigado(a) por entrar em contato / contactar com o suporte		Comunicações		
the <b>transfer</b> will take [time] <b>to reach the recipient account.</b>	a transferência vai levar [tempo] para cair na conta de destino / para chegar à conta de destino	a transferência vai levar [tempo] para chegar à conta de destino		Finanças	<a href="https://transferwise.com/br/blog/quanto-tempo-demora-para-cair-uma-transferencia-bancaria">https://transferwise.com/br/blog/quanto-tempo-demora-para-cair-uma-transferencia-bancaria</a>	
This is why (COUNTRY - MALE) is not <b>showing in the drop down</b> of	É por isso que o (PAÍS – MASC) não está <b>aparecendo</b> na lista	É por isso que o (PAÍS – MASC) não está <b>a aparecer</b> na		TI		

countries that you can choose.	de países disponíveis.	lista de países disponíveis.				
This should <b>take care of the problem</b> .	Isto deve tratar do problema	Isto deve resolver o problema		língua comum	Linguee, Dicio	
To verify it, please click the <b>link to verify your email</b> in the email verification email.	por favor, clique no link de verificação de e-mail, no e-mail de verificação.			TI		
top-level domains (website confirmation)	domínios de nível superior			TI	Microsoft Terminology Collection	
transfers details	dados da transferência			finanças		
travel provider				Transporte		
troubleshooting tips	dicas de solução de problemas	sugestões para resolução de problemas	Passos para o processo de diagnosticar a origem de um problema.	TI	Microsoft Terminology Collection, <a href="https://techterms.com/definition/troubleshooting">https://techterms.com/definition/troubleshooting</a>	
turn around time	prazo	prazo de execução	Período decorrido entre a entrega ou início de um trabalho a realizar e a obtenção dos resultados (por exemplo, no caso de análises laboratoriais, é o prazo de entrega dos resultados).	Transporte	IATE, Linguee	

tv shows and episodes	programa de televisão, série de televisão, série televisiva, série de TV, telessérie		Um tipo de programa televisivo ou programa online com um número pré-definido de capítulos por temporada, chamados episódios.	Entretenimento	Linguee, Wikipedia	
unsubscribe / cancel subscription	cancelar a assinatura / inscrição	cancelar subscrição	Remover-se como destinatário de uma lista de endereçamento.	TI	Microsoft Terminology Collection	
upload documents	Subir / fazer o upload do documento	Carregar o documento		TI	Microsoft Terminology Collection	
Upload documents section	seção "fazer o upload dos / subir documentos"	secção "carregar documentos"	Um controle que contém a função de enviar dados de um computador local para um computador ou servidor remoto, geralmente através da internet.	TI	Microsoft Terminology Collection, <a href="https://www.softonic.com.br/artigos/mega-como-subir-um-arquivo-sem-complicacoes">https://www.softonic.com.br/artigos/mega-como-subir-um-arquivo-sem-complicacoes</a>	No glossário da Microsoft também há “carregar documentos” para PB.
upload file / upload the file	Subir / fazer o upload do arquivo	Carregar o ficheiro	Ação de enviar dados de um computador local para um computador ou servidor remoto, geralmente através da internet.	TI	Microsoft Terminology Collection, <a href="https://www.softonic.com.br/artigos/mega-como-subir-um-arquivo-sem-complicacoes">https://www.softonic.com.br/artigos/mega-como-subir-um-arquivo-sem-complicacoes</a>	No glossário da Microsoft também há “carregar arquivo” para PB.

user identification	identificação do usuário	identificação do utilizador	Entidade lógica (constituída geralmente por uma palavra ou uma combinação alfabética ou alfanumérica) utilizada para identificar o usuário de um programa, sistema informático ou serviço de acesso à Internet e que o distingue dos demais usuários.	TI e processamento de dados	Microsoft Terminology Collection, IATE	
utility bill	conta	conta / fatura de serviços públicos	Documento que solicita o pagamento a ser feito para empresas de uma jurisdição local. Estas contas exigem pagamento por um serviço público prestado e recebido pelos ocupantes de uma casa.	Finanças	Linguee, <a href="https://www.reference.com/business-finance/utility-bills-b7a3cd91f562e028">https://www.reference.com/business-finance/utility-bills-b7a3cd91f562e028</a>	
verification request	solicitação de verificação	pedido de verificação		TI	Microsoft Terminology Collection	
video call	chamada de vídeo / videochamada		Qualquer chamada telefônica que inclua um componente de vídeo, onde você possa ver a pessoa com quem está falando (e eles podem vê-lo).	TI	Microsoft Terminology Collection	
we are Cc'ing the login email	estamos enviando com cópia para o e-mail de logon	estamos a enviar com cópia para o endereço de correio eletrónico do início de sessão	Enviar (uma cópia de um documento, e-mail ou algo semelhante) para alguém.	TI	<a href="https://www.dictionary.com/browse/cc-ing">https://www.dictionary.com/browse/cc-ing</a>	cc (com cópia) em português

we are <b>putting increasing resource</b> into monitoring	estamos aumentando recursos em monitoramento / estamos colocando cada vez mais recursos...	estamos a aumentar recursos em monitorização / estamos a colocar cada vez mais recursos...		língua comum	Linguee	
We currently	no momento	de momento		língua comum	Cyberdúvidas	
We did not receive enough information on your previous <b>application</b> and due to this it was <b>closed</b> .	Não recebemos informações suficientes sobre o <b>seu pedido</b> / <b>processo</b> anterior e por isso ele foi <b>arquivado</b>			TI, suporte técnico	Linguee	Não confundir com “candidatura” ou “aplicativo”.
We will be glad / happy (to help, to answer), We're happy to be of assistance.	Será um prazer (ajudá-lo).	teremos muito gosto (em ajudar)		Saudações profissionais	Linguee, <a href="https://ciberduvidas.iscte-iul.pt/consultorio/perguntas/ter-muito-gosto-em-vs-ter-o-gosto-de/34816">https://ciberduvidas.iscte-iul.pt/consultorio/perguntas/ter-muito-gosto-em-vs-ter-o-gosto-de/34816</a> , Linguee	
we're working hard	estamos trabalhando muito / arduamente	estamos a trabalhar muito / arduamente		língua comum	Linguee	
which I appreciate	agradeço		Ser grato/a por algo.	língua comum	Collins Dictionary	
wholesale rate	taxa por atacado	tarifa por grosso		Finanças	Linguee	

wire transfer	transferência eletrônica / bancária	transferência bancária. (Menos frequente: transferência eletrônica)	Um método de transferência de recursos financeiros de uma conta bancária para outra.	Finanças	Linguee, Collins Dictionary	
withdraw cash	sacar / retirar (dinheiro)	levantar (dinheiro)	Retirada de dinheiro feita em qualquer agência bancária.	Finanças	<a href="https://www.bb.com.br/pbb/pagina-inicial/voce/produtos-e-servicos/contas/todos-os-servicos/saque-movel">https://www.bb.com.br/pbb/pagina-inicial/voce/produtos-e-servicos/contas/todos-os-servicos/saque-movel</a> , IATE, Linguee, Dicionário Informal	
withdrawing funds	saque / retirada de fundos	levantamento de fundos	Retirada de fundos de investimento.	Finanças	<a href="https://www.bb.com.br/pbb/pagina-inicial/voce/produtos-e-servicos/contas/todos-os-servicos/saque-movel">https://www.bb.com.br/pbb/pagina-inicial/voce/produtos-e-servicos/contas/todos-os-servicos/saque-movel</a> , IATE, Linguee	
within the hour	em uma hora / dentro de uma hora	dentro da hora / de uma hora	Antes que muito tempo tenha passado, geralmente menos de uma hora.	língua comum	<a href="https://www.macmillandictionary.com/dictionary/british/within-the-hour">https://www.macmillandictionary.com/dictionary/british/within-the-hour</a> , Microsoft Terminology Collection, Linguee	Não confundir com “na hora”, que significa “imediatamente”.
you can try heading here (head for)	pode tentar vir aqui		direcionar o caminho de alguém (para)	língua comum	Collins, Linguee	Não confundir com “heading” (título).
you have <b>locked yourself out of the application</b>	o aplicativo foi bloqueado / o seu acesso ao aplicativo foi bloqueado	a aplicação foi bloqueada		TI	Linguee, Microsoft Language Collection	